THE
PROCEEDINGS
OF
THE PHYSICAL SOCIETY
Section B

CONTENTS

PROCEEDINGS OF THE PHYSICAL SOCIETY

The *Proceedings* is now published monthly in two Sections.

ADVISORY BOARD

Papers for publication in the *Proceedings* should be addressed to the Hon. Papers Secretary, Dr. H. H. Hopkins, at the Office of the Physical Society, 1 Lowther Gardens, Prince Consort Road, London S.W.7. Telephone : KENsington 0048, 0049.

Detailed Instructions to Authors were included in the February 1948 issue of the *Proceedings* ; separate copies can be obtained from the Secretary-Editor.

BULLETIN ANALYTIQUE

Publication of the Centre National de la Recherche Scientifique, France

The *Bulletin Analytique* is an abstracting journal which appears monthly in two parts, Part I covering scientific and technical papers in the mathematical, chemical and physical sciences and their applications, Part II the biological sciences.

The *Bulletin*, which started on a modest scale in 1940 with an average of 10,000 abstracts per part, now averages 35 to 40,000 abstracts per part. The abstracts summarize briefly papers in scientific and technical periodicals received in Paris from all over the world and cover the majority of the more important journals in the world scientific press. The scope of the *Bulletin* is constantly being enlarged to include a wider selection of periodicals.

The *Bulletin* thus provides a valuable reference book both for the laboratory and for the individual research worker who wishes to keep in touch with advances in subjects bordering on his own.

A specially interesting feature of the *Bulletin* is the microfilm service. A microfilm is made of each article as it is abstracted and negative microfilm copies or prints from microfilm can be purchased from the editors.

The subscription rates for Great Britain are 4,000 frs. (£5) per annum for each part. Subscriptions can also be taken out to individual sections of the *Bulletin* as follows :

	frs.	
Pure and Applied Mathematics—Mathematics—Mechanics	550	14/6
Astronomy—Astrophysics—Geophysics	700	18/–
General Physics—Thermodynamics—Heat-Optics—Electricity and Magnetism	900	22/6
Atomic Physics—Structure of Matter	325	8/6
General Chemistry—Physical Chemistry	325	8/6
Inorganic Chemistry — Organic Chemistry — Applied Chemistry—Metallurgy	1,800	45/–
Engineering Sciences	1,200	30/–
Mineralogy—Petrography—Geology—Paleontology	550	14/6
Biochemistry—Biophysics—Pharmacology	900	22/6
Microbiology—Virus and Phages	600	15/6
Animal Biology—Genetics—Plant Biology	1,800	45/–
Agriculture—Nutrition and the Food Industries	550	14/6

Subscriptions can be paid directly to the editors : Centre National de la Recherche Scientifique, 18, rue Pierre-Curie, Paris 5ème. (Compte-chèque-postal 2,500–42, Paris), or through Messrs. H. K. Lewis & Co. Ltd., 136, Gower Street, London W.C. 1.

The Physical Basis of Life*

BY J. D. BERNAL†

Birkbeck College, London

Mr. Chairman, Ladies and Gentlemen,

I feel deeply honoured at being asked to join the very distinguished series of Guthrie lecturers. I have chosen a subject which at first sight seems rather far removed from the work of Professor Guthrie. However, physics has changed much since his time, and even then the foundations were being laid for its transition towards biology. I only know two persons who were students of Professor Guthrie, men of almost completely contrasting temperaments, H. G. Wells and Professor Tutton. The first, as a man of imagination and of qualitative views, did not take kindly to the intensely practical nature of Professor Guthrie's physics course, and has left an amusing account of it in his *Experiment in Autobiography*—
" . . . . Now when I came into the physics laboratory I was given a blowpipe, a piece of glass tubing, a slab of wood which required planing and some bits of paper and brass, and I was told I had to make a barometer. So instead of a student I became an amateur glass worker and carpenter.

" After breaking a fair amount of glass and burning my fingers severely several times, I succeeded in sealing a yard's length tube, bending it, opening out the other end, tacking it on to the plank, filling it with mercury, attaching a scale to it and producing the most inelegant and untruthful barometer the world has ever seen. In the course of some days of heated and uncongenial effort, I had learnt

---

* The title of this lecture is an unconscious echo of the once celebrated lecture of T. H. Huxley (1901) delivered in Edinburgh in 1868. I did not read that lecture till long after I gave this one, but it is interesting to note the changes that eighty years have brought since it was first clearly stated that life had *one* physical basis. Huxley had a word for it—Protoplasm. He was concerned with stressing that it was functionally, formally and substantially the same over the whole range of living things. In function all organisms showed metabolism, movement, growth and reproduction; in form they were all composed of nucleated cells; in substance they were all made of protein, a compound of carbon, hydrogen, oxygen and nitrogen. Not only was life one, but it was linked materially with the non-living, by the capacity of converting inorganic gases and minerals into protoplasm.

The latter half of the lecture was spent in defending the thesis of the former from the charge of rank materialism on the grounds that we did not know any more what matter was than what spirit was, and it was reasonable to accept our ultimate ignorance, however laudable our efforts at understanding. This was the basis of his famous attitude of *Agnosticism*.

The positive part of his lecture stands as firm today as it did when he first put it forward. What is curious, however, is that he nowhere attempted, great evolutionist as he was, to explain the unity of all life in terms of a common history. The developments of physiology and biochemistry have filled in the details of Huxley's picture but they add nothing fundamentally new to it unless they leave the field of describing structures and mechanism and enter into that of searching for origins. It is here, if anywhere, that lies the excuse for making a new attempt at understanding the physical basis of life.

† This article also appears in the September issue of *Proc. Phys. Soc.* A (p. 537).

nothing about the barometer, atmospheric pressure, or the science of physics that I had not known thoroughly before I left Midhurst, unless it was the blistering truth that glass can still be intensely hot after it has ceased to glow red.

" I was then given a slip of glass on which to etch a millimetre scale with fluorine. Never had millimetre intervals greater individuality than I gave to mine. Again I added nothing to my knowledge—and I stained my only pair of trousers badly with acid."

Professor Tutton, on the other hand, one of the most precise of physicists and crystallographers, rather overdid it in the other direction. Apparently the barometer he produced was not only perfect in every detail but was varnished to boot, and legend has it that Professor Guthrie proceeded to smash it up on the grounds that he was training physicists and not instrument makers.

In one particular respect Professor Guthrie's work does touch the subject of this lecture. He was one of the first to examine the properties of the cryohydrates, those salts in which a large amount of water is held in fixed positions similar to those they occupy in ice. As I shall show, cryohydrates are a significant feature in the crystallography of the proteins and may have had an important part to play in the origin of large molecules adsorbed on clay particles.

I have chosen a very general subject for this lecture mainly because, owing to the events of recent years, I have been unemployed as an experimental physicist for longer than I like to think, and have not been long enough back at work to be able to produce new material on the detailed study of any physical field. Instead, I want to discuss in a general way one of the aspects of physics that is coming more and more into prominence at this time—its relation to the processes of life. At present there are two really exciting fields in physics: one of them is at the outer edge of physics—nuclear physics, and the even more short-period and energetic processes of cosmic rays—and the other at the inner edge—where physics touches biology, a field in which the interest lies in complexity rather than in intrinsic energy. In saying this I am treating physics in its modern and more extensive connotation as including chemistry, because, although chemical manipulation and the logical structural analyses of the older chemistry still remain, and the direct study of material substances and transformations will continue as a source of knowledge, chemistry and physics are now embedded in one common theory and are in future bound to become one integral discipline. Accordingly biophysics and biochemistry, to make any sense, must be considered together.

What I am trying to put before you is the first crude attempt at stating the problems of the origin and function of life from a physical standpoint. In doing so I am fully aware of my inadequacy. It is probable that even a formulation of this problem is beyond the reach of any one scientist, for such a scientist would have to be at the same time a competent mathematician, physicist, and experienced organic chemist, he should have a very extensive knowledge of geology, geophysics and geochemistry and, besides all this, be absolutely at home in all biological disciplines. Sooner or later this task will have to be given to groups representing all these faculties and working closely together theoretically as well as experimentally. The most I can do is to indicate some of the fields in which the key problems exist and provide some preliminary guesses as to the direction in which to look for solutions.

The approach I am choosing is the broadest one possible, and deliberately so. I am not here concerned with the task of elucidating this or that physical structure

or mechanism underlying particular functions of existing organisms, but rather with the whole range of the phenomena which we commonly call life. But my emphasis will be on one particular aspect of life, on the problem of *origin* rather than that of structure, metabolism and behaviour. I have chosen this emphasis on origin in biological systems because it is far more important than in physical systems. Until recently discussions on the origins of systems were considered in some way improper to science, but now even in physics itself the questions of origins are coming into discussion, as, for example, those of the origin of the nebulae, the solar system and the elements themselves. Most of what I say will not be new, it is largely based on the writings of other workers, notably Haldane (1928), Lwoff (1943), Oparin (1938) and Dauvillier (1947). What I have done is to correlate these different contributions and add some speculations of my own on the actual conditions under which life originated. My aim is by such a broad sketch to bring out in sharp relief the critical points of difficulty, not in order to evade them by pious allusions to mysteries beyond human comprehension, but as a guide to practical research in the future.

The first obvious step in the discussion should be to define the terms. But life is an extremely difficult term to define. In fact Dr. Pirie (1937) in his paper on " The meaninglessness of the terms ' Life ' and ' Living '" has practically proved that it is incapable of definition:

" We have now examined destructively the various qualities which might be used to define the word ' life ' and we have found that they are individually inadequate for even an approximate definition. There is not space here to discuss what might be done with all the permutations of these qualities ; it may be said however, that combinations of two or three qualities, though they might easily be drawn up to exclude all obviously non-living systems, will also exclude some which are, if not typically living, at least generally included in that category.

" Until a valid definition has been framed it seems prudent to avoid the use of the word ' life ' in any discussion about border-line systems and to refrain from saying that certain observations on a system have proved that it is or is not ' alive '."

By taking every quality, such as respiration or movement, usually considered characteristic of life, he is able to show that there are many things called " alive " which do not possess any of them, and equally many possessing some of them that we do not call " alive ".

I wish to avoid these troubles by not attempting to give a definition of any particular form of life or living organism, but rather by treating the common word " life " as referring particularly to the totality of processes accessible to our observation on this earth : to what Goldschmidt has called the biosphere, consisting materially of the group of complex organic compounds found almost exclusively in the watery layer on the surface of the earth, the hydrosphere, or in the adjoining regions of the atmosphere or parts of the lithosphere clearly derived from it. If we limit it in this way we can for the moment find one common material characteristic, the presence of protein molecules, and at the same time one common physico-chemical process, the stepwise catalysis of organic compounds carried out practically isothermally by quantum jumps of between 3 and 16 kilocalories, small compared with the usual jumps of 300 in laboratory chemistry.

These terrestrial limitations obviously beg the question of whether there is any more generalized activity that we can call life. Biology in this respect is on a different basis from physics and chemistry in that it deals less with universals

and more with contingents. It belongs to the kind of descriptive and inter-pretative studies we might more properly call " graphies ", including observational astronomy and geography. Whether there are some general characteristics which would apply not only to life on this planet with its very special set of physical conditions, but to life of any kind, is an interesting but so far purely theoretical question. I once discussed it with Einstein, and he concluded that any generalized description of life would have to include many things that we only call life in a somewhat poetical fashion. Any self-subsisting and dynamically stable entity transforming energy from any source, or, as Haldane put it, " any self-perpetuating pattern of chemical reactions ", might be called " alive " in this sense. The value of distinguishing it as an individual system or organism would only exist if the total phenomena persisted for a time appreciably longer than the periods or characteristic times of any internal processes it might contain. In this sense a galaxy or a star is alive, or, on a more terrestrial scale, a flame. Passing to a degree of complication greater than the biological, we might talk of the life of a human culture or civilization. All are characterized by birth, persistence and death.

These considerations, however, are all very abstract. Their only immediate value is to indicate that though there may be a number of kinds of life in this sense, these kinds of life are quite distinct, and each has a very limited sphere of existence. All must satisfy two necessary criteria. The first is that they must be functional, in the sense that the processes that comprise them must have a certain dynamic stability.*

Now this leads us back again to the method of tackling the problem as one at the same time historical and physiological. The actual material structures which build up our terrestrial life, with which we shall be exclusively concerned from now on, are to be considered at the same time under both aspects, those of their function and of their origin. An object such as a galaxy or an apple carries its whole significance only if we consider the stages of its development as well as its instantaneous activity. The ostensible purpose of biology in unravelling the processes that occur in living things is at the same time the elucidation of the necessary stages by which they arrived at their actual structures. Present study throws light on past history and vice versa. Every existing organism is in this sense a fossil. It carries in it by inference all the evidence of its predecessors;† and this remains the case even if we cannot read it clearly or at all. Naturally the study of the past, which cannot be directly determined, can provide no positive proof to set against immediate observation of the present ; but it can do something equally important. It can give indications as to where to look in the present to find significant things.

The history of science affords many remarkable examples of this process. The chief of them of course is Darwin's statement of natural selection as a mechan-ism for evolution. Undoubtedly evolution is a hypothesis, natural selection is

---

* The meaning of this first criterion has been made much clearer by the work of Prigogine (1947) on the thermodynamics of " open systems ". Such a system can take in and pass out matter as well as energy and is not subject to the second law of thermodynamics, that is, its entropy may remain constant or even decrease, the condition of stability of such a system being that the rate of change of entropy is a minimum; and secondly, they must originate and develop out of some pre-existing system, or, in plain English, they must work and they must have got there in the first place.

† Professor Haldane says " No, it doesn't. Gene segregation is a trick for halving the amount of the ' evidence ' once per metazoan life cycle. It may carry enough evidence." It would have been better perhaps to have phrased the statement conversely as " everything an organism contains is evidence of its predecessors ".

an observational fact : but it was from the hypothesis of evolution, that is, from the need to provide an explanation of present forms of life in terms of older ones, that Darwin was driven to use the natural selection argument which he borrowed from Malthus. As a result he not only put into order the chaos of the accumulating biological observations of his time, but led to the fertile and useful experimental studies of ecology, morphology and heredity.

Darwin himself, curiously enough, was opposed to any generalization of his own methods. In a letter to Sir J. Hooker (see Darwin 1887) he said : "It is mere rubbish thinking at present of the origin of life ; one might as well think of the origin of matter." We are now almost in a position to take him at his word. For the origin of matter, at least the origin of the forms in which matter presents itself on this earth, is at last becoming clear. The chemical elements, in their ordered multiplicity, and with their apparently arbitrary and enormous variations of abundance, are seen now as perfectly logical products of a process going on in some primordial hyperstar or concentrated universe some four billion years ago. In fact, the physicists have reconstructed, somewhat provisionally, the temperatures and pressures of that time from the relative abundances of different elements, and in the process have discovered anomalies which would not have been discovered without some such hypothesis as to origin.

This does not mean that we should accept wild hypotheses of the origin of life or of matter which simply conceal ignorance, but rather that we should attempt almost from the outset to produce careful and logical sequences in which we can hope to demonstrate that certain stages must have preceded certain others, and from these partial sequences gradually build up one coherent history. There are bound to be gaps where this cannot be done, but until the process is attempted these gaps cannot be located, nor can the attempt be made to fill them. The process is not dissimilar, though on a vastly greater scale, to the attempts of the geologists two hundred years ago to determine the history of the crust of our globe, which, once theological preoccupations had been overcome, led to the ordered logical sequences of Hutton and Lyell.

There are a certain number of general considerations which we can employ in building these sequences. In the first place we may know that we can expect gaps where processes are particularly rapid. In any sequence which passes through a number of phases the occurrence of any phase will be proportional to its duration. Unstable elements are the least abundant. Rapidly evolving forms leave the fewest fossils. What we observe easily are the stable stages, and they may form a very small fraction of the whole process, giving a much greater impression of discontinuity than really exists. Life as we know it consists of only a relatively small number of kinds of simple molecules—sugars, amino acids, purines-— molecules out of which are composed the far more complicated macro-molecules of the proteins and nucleic acids and microscopically visible structures such as membranes and fibres. We now know, particularly through the use of tracer elements, that there exist as well a large number of intermediate molecules which are changed so quickly that they cannot be observed by ordinary methods. But we may further infer that in the history of life there have been formed a still larger number of molecules that have played a decisive rôle in its chemical evolution, but which can be reconstructed now only from the traces they have left in the structure of existing molecules.

Another very general consideration for all developing systems is the way in which processes occurring for the first time react on those already established. One of the perennial arguments in biology has been the question of spontaneous generation. Ceaselessly put forward to explain otherwise obscure phenomena, it has always been disproved whenever careful work has been carried out, and we are apt to think that Pasteur has settled it for all time. But those who demonstrated that spontaneous generation did not normally occur under existing conditions usually went further and claimed that they had also proved that it could never have occurred at all by natural means, and, therefore, only substituted an original miracle for a series of continuing ones (see Engels 1940). Now one obvious weakness of this argument is that the conditions in which life originated do not exist any more on this earth. It is not necessary, to prove this, to assume any gross change in the external physical conditions, surface temperature or solar energy, etc. It is sufficient to know, as Haldane (1928) has pointed out, that the very existence of life itself radically changed the initial conditions, principally by producing molecular oxygen. Similarly, at almost every stage in the evolution of life, successful new forms interfered with all that had gone before. There is no guarantee, of course, that life itself will not so interfere with the universe into which it has intruded itself as to wipe itself out. All it has done so far is to block possibilities of development other than those actually taken. This implies, of course, a potentiality for life much more extensive than the actuality.

There is a definite and limited range of potentiality fixed at a lower level of complication by the physical and chemical properties of molecules. This corresponds to what Henderson spoke of in his remarkable book of forty years ago, *The Fitness of the Environment*. He pointed out that the peculiar properties of water and carbon dioxide, for instance, were as suitable for life as life was adapted to them. But though life in its terrestrial form depended on these properties, we now see that this is about equivalent to saying that life depends on the laws of arithmetic or geometry, because the peculiar properties of water and carbon dioxide are implicit in the number of electrons contained in their atoms, and that the abundance of atoms with these numbers and their presence on the surface of the earth follows logically from their nuclear composition. When these interactions, however, become as complicated as they have done in life, we can no longer be sure that the only absolutely possible track has been taken. Indeed, from the very variety of life on this earth, in past as in present times, we know what a great range of possibility there has been and what apparently small accidents, some of them outside the range of life altogether, such as those of geography or climate, have determined the success or failure of this or that form.

We can if we like call these possibilities the Aristotelian entelechies, and this can be done without any mystical invocations. Any arrangement of atoms or molecules necessarily carries with it complex possibilities of order and function, and those are *immanent* in the structure of the molecules themselves. Which appear and which do not appear are, however, immanent not in any particular combinations themselves but in the total play of universal forces, and, therefore, for the consideration of the system, may be considered as *contingent* or accidental. It is a convenience for us to separate the immanent and contingent elements, attributing the first to science and the second to cosmic history, but their interaction must always be taken into account. Deepening our knowledge of the behaviour of material systems helps us to detect and to understand historical

changes.   Conversely, knowledge of the origin and history of life gives a clue to the understanding of biochemistry and microbiology.

The new interest in the physical nature of biological systems has coincided necessarily with the development of new physical tools, theoretical and practical. Of the former, by far the most important is the development of the quantum theory and its extension to cover at least in principle the theories of chemistry and to give precious indications of the far more complex phenomena of biochemistry and biophysics.   Without exaggeration it can be said that the conception of quantum energy changes in chemical reactions is the most illuminating and the most effective new idea in modern biology.   We begin to see now that the material aspects of a living system are but the struts and levers of a machine, the particular function of which is to effect energy interchanges, and that growth and assimilation are but means of achieving a metabolism consisting of enzyme promoted energy changes.   Thus in a very physical sense process takes precedence over structure. The developments of other aspects of physical science also have theoretical applications to biology, particularly the theory of solids and liquids, which may help to explain the apparently mysterious activities which go on inside living cells. On the other side our new knowledge of soil science and of geochemistry, as well as that of physical geography and oceanography, throw light on the medium from which life evolved and which still supports it.

With these advances in understanding has come a much greater armoury of experimental methods.   On every level of organization, both of structure and of energy change, new physical and chemical instruments are offering the biologist a range of opportunity unequalled since the first use of the microscope and of the balance in the seventeenth century.   On the chemical level the introduction of radioactive tracers is likely so to revolutionize the unravelling of reactions in both organic and biochemistry as to make what we think we know now obsolete in a matter of a few years.   For the determination of molecular structure, the use of spectroscopy, particularly infra-red spectroscopy, and of x-ray crystallography combined, gives us a picture of atomic arrangements limited in its accuracy and extent only by the number of workers in the field and the slowness of the computations involved ; and this limitation itself is likely to be removed in a few years by the further development of electronic calculating machines which will enable some years' work to be done in a few hours.   Chemical methods themselves have enormously increased their range and accuracy, owing to the development of industrial polymer chemistry.   Differential chromatography, which enables the simple and complex constituents of proteins and other organic substances to be analysed and labelled with ease and certainty, provides a bridge between the chemistry and the physics of large molecules.

It is, however, in the intermediate zone, the zone between the chemical molecule of some tens of Ångströms in dimensions, and the old limit of microscopy at 2000 Ångströms, that the most significant advances have been made.   Indirect methods of great beauty and interest we have had for some time—the use of the ultra-centrifuge by Svedberg, of electrophoresis by Tiselius, of viscosity by Staudinger, and more recently of light scattering by Doty.   These provided the first really quantitative picture of the approximate size and shape of the large molecules of biological and industrial origin.   Now, however, we have far more precise and accurate methods.   X-ray analysis can be extended to deal with molecules of hundreds of Ångströms in dimensions, and although it cannot as

yet give the precise details of their atomic constitution, it has already led to significant information as to their general nature, particularly of their constancy under different physical conditions.

Far more spectacular in recent years, however, has been the development of the electron microscope, making visible in an immediately understandable form structures from those containing a few score atoms to the limits of microscopic vision and beyond. We are certainly now in a Galilean phase of observational biology. Until the advent of the electron microscope, a great blank existed between the knowledge of atomic combinations provided by chemistry and that of histological structures observable with the microscope. This gap was filled with the mystic word "colloid", which served to explain the very real but very obscure properties depending on the existence of structures of magnitudes between ten and ten thousand Ångströms. Now the colloid world is open for inspection, and the term itself will probably vanish or acquire a precise and limited meaning. The coming of the electron microscope, as has often happened in the history of science, has stirred the optical microscopists into exertions which their sole possession of the field did not provoke. The microscope had, in fact, stood virtually still between 1880 and 1940, but now, with the rival in the field, it was found that it could be improved in many ways, and the new phase microscopes, and the ultra-violet, infra-red, polarizing and reflecting microscopes provide an armoury which, though they cannot rival the electron microscope in resolving power, have, in the hands of such workers as Casperssen, shown a power of chemical interpretation of structure greater than any other method, and having the additional enormous advantage that they can be used on living material. With all these new methods simultaneously available it is not surprising that the present picture of biological research should be one of exciting but somewhat disordered advance— an advance so rapid that attempts to assess the position will be out of date almost as soon as, if not before, they are written.

In the account which follows an attempt is made to present the main outlines and the critical stages in the development of life from its inorganic origins. It is based essentially on two kinds of data—the geochemistry and physico-chemistry of the cooling planet, and the organic chemical composition common to all existing living organisms. Such an attempt reveals at once the large gaps that still exist, but it also reveals the lack of perfectly feasible research which is bound to help to reduce these gaps and to bring out others that may now be unsuspected.

The process is one which we can imagine as taking the form of a play divided into a prologue and three acts. The prologue introduces the scene on the surface of the primitive earth, and the first group of actors of an entirely inorganic kind which must start the play. The first act deals with the accumulation of chemical substances and the appearance of a stable process of conversion between them, which we call life ; the second with the almost equally important stabilization of that process and its freeing from energy dependence on anything but sunlight. It is a stage of further synthesis and of the appearance of molecular oxygen and respiration. The third act is that of the development of specific organisms, cells, animals and plants from these beginnings. All we have hitherto studied in biology is really summed up in the last few lines of this act, and from this and the stage set we have to infer the rest of the play.

Our knowledge of the nature of the surface of the primitive earth is derived from the spectroscopic evidence as to the abundance of the elements in the stars, and the much more meagre information we have about them in the atmospheres

of the planets on the one hand, and on the other the knowledge of geochemistry and geophysics. This field of study has been entirely transformed in recent years through the application of x-ray crystallography in the hands particularly of the Braggs and Goldschmidt. In the latter's great work, *Verteilungsgesetz der Elemente* (1923–37), are traced the processes by which the primitively mixed elements of the original solar filament sorted themselves out in the first place under simple gravity, and at the latter stages by crystallization according to the laws of ionic combination, which he discovered and Pauling later refined. It is still an open question whether the first sorting process was ever complete, whether the still liquid interior of the earth consists of iron or of unconsolidated solar matter, mostly of hydrogen, as Kuhn (1946) has recently proposed.

The outer layers, however, in their melted form, must have contained in addition to silicates—predominantly basic silicates—much water and carbonate in solution, the whole being originally surrounded by an atmosphere of hydrogen and hydrides, CH_4, NH_3, H_2S, and H_2O though in far less quantities than now. This original atmosphere must have been modified in two particular ways. The actual crystallization of the crust must have forced into the atmosphere vast quantities of water vapour and carbon dioxide. At the same time hydrogen was being steadily lost at the top of the atmosphere, the gravitational pull not being strong enough to hold it. The result is bound to have been a steady oxidation, methane turning to CO_2 and water through intermediate compounds such as aldehydes, alcohols and acids which may have had a rôle in the formation of life ; ammonia oxidized to nitrogen and H_2S to sulphur. Further oxidation seems most unlikely at the pre-organic stage. Some must have occurred at the very top of the atmosphere in the dissociation of water to hydroxyl and free hydrogen, but the small amount of hydroxyl formed would be used up on oxidizing compounds in the lower atmosphere, and there remained an enormous residue of reducing material in the ferrous iron of the then exposed primitive rocks.

By the time the earth had cooled sufficiently for the water to condense, the atmosphere may have been largely one of nitrogen with a gradually decreasing concentration of carbon dioxide as its partition coefficient in the cooling sea became less and less. The seas would contain primarily ammonia, carbon dioxide and hydrogen sulphide in solution. Whether they also contained salt in anything like the present concentration is still an open question, though there are enough minerals containing chloride to provide for present concentration if sufficient rock had been worked over in the course of geological history by weathering processes ; but on the other hand it is possible that vapourized halides formed parts of the original atmosphere, came down molten, solidified and were dissolved in the primitive seas (Dauvillier 1947).

The surface of the world at this stage must not have been very different from what it is now, except for the bareness of the rocks and consequently greater speed of weathering. We still do not understand the mechanism of continent and mountain formation, but there is no reason to believe that it was dependent on organic processes. We may therefore assume most of the geographical features which we now observe, with the exception of coral islands, though inorganic calcium carbonate precipitation may have occurred. In particular there must have been, as now, extensive areas of mud in deltas and on continental shelves, some of which would be exposed to the tides. Correspondingly on the land a kind of soil would be formed wherever the run-off was not sufficient to remove weathered material or where it was deposited in rivers, and this soil would also contain clay.

At first sight there might be no reason why the world should not have continued in this state indefinitely, but there was an active agent operating at that time which is no longer in operation, namely the influx of solar radiation out to the far ultra-violet of 2000 A. or less. If we put to a physical chemist the problem of the reactions occurring in a weak solution of ammonium carbonate and sulphide under such radiation, he would agree that although it was not possible without experiment to determine exactly what compounds would be formed, a process of polymerization and condensation leading to the formation of nitrogenous organic compounds such as the amino acids is almost certain, and would proceed until an equilibrium was reached where breakdown was equal to formation. It would be of the utmost importance to make these vague remarks more precise, and researches in ultra-violet photosynthesis would certainly be of interest, and might be of practical use.

Equilibrium, moreover, could be reached in two ways. One would be the straight photochemical process of absorption and emission ; the other might well be a dark breakdown through a number of intermediate compounds liberating energies corresponding to smaller quantum steps and longer wavelengths. Life from the purely physico-chemical point of view is simply a denotation of the complex mechanism of this latter process. We may give a schematic picture of this in the form of a graph showing purely qualitatively the effective utilization of sunlight at different periods of world history. At first the absorption is determined only by the original constituents of CO_2 and ammonia. Later the more complex forms give rise to greater absorption. When life appears, and these forms are broken down, the absorption decreases. The subsequent hypothetical changes due to the appearance of organic photosynthesis and respiration will be discussed later.

The stage is now set for the appearance of life itself. In Moleschott's classic phase, "It is woven out of air by light". It is here that organic chemistry first begins. Condensations and dehydrogenations are bound to lead to increasingly unsaturated substances, and ultimately to simple and possibly even to condensed ring structures, almost certainly containing nitrogen, such as the pyrimidines and purines. The appearance of such molecules makes possible still further syntheses. The primary difficulty, however, of imagining processes going thus far is the extreme dilution of the system if it is supposed to take place in the free ocean. The concentration of products is an absolute necessity for any further evolution. One method of concentration would of course take place in lagoons and pools which are bound to have fringed all early coastlines, produced by the same physical factors of wind and wave that produce them today. It has occurred to me, however, that a much more favourable condition for concentration, and one which must certainly have taken place on a very large scale, is that of adsorption in fine clay deposits, marine and fresh water. Our recent knowledge of the structures of clays has shown what an enormous rôle they still play in living processes. There is probably today more living matter that is protein in the soil and in the estuarine and sea-bed clays than above the surface or in the waters. Now the effective part of this fine-grained clay is known, particularly through the electron microscope studies of Hast (1947), to consist of what might reasonably be called clay molecules, single layers of aluminium silicate some ten Ångströms thick and a hundred and forty across, covered on both surfaces with hydroxyl groups and capable of adhering with a larger or smaller number of water molecules

into small pseudo-crystals like piles of coins. Such a small clay particle has an enormous effective adsorptive surface. It has already been shown, particularly by MacEwan (1948), that organic chemicals of a wide variety are preferentially adsorbed on such surfaces in a regular way. It is therefore certain that the primary photochemical products would be so adsorbed, and during the movement of the clay might easily be held blocked from further possibly destructive transformations. In this way relatively large concentrations of molecules could be formed.

This formation is impossible nowadays for two reasons: firstly the cutting off of the ultra-violet by the ozone layer, and secondly the almost universal presence of life, which would destroy such molecules if they were formed in any other way. The very absence of life ensures an accumulation of material containing available energy for indefinite periods : the original world was sterile. Now the clay molecules have another property besides adsorption. Small molecules attached to them are not fixed at random, but in definite positions relative not only to the clay but to each other, and held in such positions they can interact and form more complex compounds, especially if energy can be supplied in the form of light. Clays are now one of the most important of industrial catalysts*. Polymerization would take place particularly easily with unsaturated compounds, with relatively free electrons. It is in this way that we may imagine that simpler molecular compounds could be made to undergo complex polymerization, polymerization to such an extent that the macromolecules produced might be able to persist in a colloidal form even without clay, and become catalysts or, as we should now call them, enzymes in their turn.

Clay is not the only material on which adsorption may take place. Quartz is another very active material which, as sand, would occur separately or together with clay. The importance of quartz in the formation of primitive molecules out of which life is constructed may be a crucial one in that quartz is the only common mineral possessing asymmetrical structure, some crystals having a right-handed twist and others a left. The characteristic of molecules occurring in living organisms, first brought out by Pasteur a hundred years ago, is that they are also asymmetric, and it has always been a very great difficulty to explain life originating with such molecules, as normal chemical processes produce right- and left-handed molecules with equal facility. It may be, of course, as Pasteur himself thought, that some general feature in the environment favours one rather than the other type, for example, the rotary polarization of moonlight or the magnetic moment of the earth ; but to me it seems more plausible that the particular twist was given at one time by the preferential adsorption of a pair of asymmetric molecules on quartz, and, as Mills has shown, once one asymmetric isomer was produced, even locally, it would produce a situation in which ultimately only one kind could be formed.

So far we have followed the track from the inorganic world by steps which, though they cannot be indicated in detail for the lack of necessary research, most

* The mechanism of catalysis and enzyme action is now becoming much clearer. It seems to depend to a very large extent on movement of hydrogen ions or protons for the most mobile of all chemical species. The most illuminating experiment carried out by Turkevich and Smith (1946) on phosphoric acid translocation of double bonds in butenes, using tritium as a tracer, shows that the actual process consists of a simultaneous transfer of a proton from one part of the molecule to the phosphate, and simultaneously one from the phosphate to another part of the molecule, thus effecting a proton transfer from one part of the molecule to the other, and leaving the phosphate without any net change. They point out, however, that most known crystals, either of a metallic kind or of an ionic kind, such as silicates, sulphates etc., can effect such transfers, and it seems highly probable that protein enzymes act in a similar way.

of which is quite feasible, nevertheless are not only plausible, but, with our present knowledge, inevitable. However, to get any significant information we must look at the other end of the play and try to draw deductions from the inner chemical structure of actual organisms. Now here we are met at the outset with a remarkable set of facts, which call for explanation in terms of origin. The overwhelming majority of living organisms, from the lowest bacteria to trees and men, are all built of a relatively small number (about thirty in all) of types of chemical molecules containing between four and forty atoms in each. Every chemical molecule has its origin in some previous combination, as certainly as every atom of every element. More complex forms, particularly in this or that organism, can be derived from those simple types, and the few cases which have been studied have been shown to be so derived.

From this it follows that there is only one predominating life, derived from one common chemical basis*. This is exactly the same logic that shows, for example, that all so-called Indo-European languages have a common set of root words, however much they have deviated afterwards. It would be wrong, however, to assume that the common molecules, the amino acids, the sugars and the purines which are the joint stock of existing life are necessarily the first organic chemicals, because there are to be found some aberrant bacteria : the purple and green sulphur bacteria which do not contain some of these molecules, in particular free sugar. It may be that modern life, as we may call it, represents a second stage, and we may have to reconstruct the first stage from these particularly primitive survivals. But all life, including these, contains one group of compounds of a far more complex nature which does seem to be of crucial importance, namely the proteins.

A hundred years ago Engels referred to life as the " mode of action of albuminous or protein substances ", and biochemical advances have only confirmed this dictum and made it more precise. The work of the last fifty years has shown something both of their function and of their structure. So many of the chemical reactions occurring in living systems have been shown to be catalytic processes occurring isothermally on the surface of specific proteins, referred to as enzymes, that it seems fairly safe to assume that all are of this nature and that the proteins are the necessary basis for carrying out the processes that we call life. Now although since the great work of Fischer we know that the proteins consist of various combinations of some twenty amino acids, we still do not know the precise structure of any of them. But we do know that they have a precise structure, and we have reasonable hope of determining it in the not so very distant future. It is perhaps significant that though the number of different proteins may be counted in tens of thousands, this represents an insignificant proportion of the possible combination of twenty amino acids. The most likely explanation is that certain sub-units containing the same amino acids in the same order must occur over and over again.

The work of the physico-chemists, particularly of Svedberg, has shown that active proteins exist in the form of molecules of definite molecular weight, and more recently x-ray structure analysis has shown that they are perfectly definite chemical compounds with identical molecules which persist unchanged through various grades of crystal hydration and into solution. We possess already much information as to the actual molecular arrangement, but unfortunately its full

* This was the major conclusion of T. H. Huxley's *The Physical Basis of Life*, see footnote, p. 597.

interpretation is a task for the future.   The situation at the moment is extremely similar to that of an archaeological expedition that has discovered large quantities of rock inscriptions in an unreadable script.   They may not know what the inscriptions mean, but they do know they mean something, and they may reasonably hope to decipher them.

The first hints at the deciphering come from the structures of degraded proteins, those used for structural purposes in the animals themselves, such as skin and hair, and those produced by violent chemical action, such as in boiling an egg.   Here Astbury (1933) has shown that polymerization must occur to form long chains, and that these chains may be straight or curved or folded in a definite plane.   As far as the straight chains are concerned this has been confirmed by the artificial polypeptides synthesized by Woodward giving similar x-ray pictures to the natural ones.   More recent work by Perutz (1949), Crowfoot (1941) and their pupils has shown that a similar arrangement must hold in the crystalline proteins, and there seems to be emerging a general picture of the primitive protein molecule, such as that derived by Crowfoot for gramicidin–s, or by Kendrew for myoglobin, indicating the existence of parallel groups of chains making up flat layers of about ten Ångströms thick and from thirty to sixty Ångströms wide.   The primitive protein may therefore have a resemblance which is more than coincidental to that of the clay particle.   More complex proteins are built by agglutination of the simpler groups.   Haemoglobin, for instance, studied in detail by Perutz, seems to consist of four of the myoglobin layers held firmly together.   In the crystal the molecules may further be separated by layers of water, just as in the clay particles.   The complication goes further with the building up of even larger protein molecules, right up to the size of the viruses whose internally crystalline nature, identical in the solid gel and solution, I was able to establish ten years ago with Bawden, Pirie and Fankuchen (1936).

Now the electron microscope has shown the absolute continuity between the external structure of protein molecules and viruses.   I am stating this without any idea of claiming that viruses are a primitive form of life.   All the evidence points in the opposite direction.   The activity of viruses is absolutely dependent on that of other organisms, and they are probably some of the most complex and sophisticated forms of life, though they seem to lack all the structures and most of the enzymes that normal organisms possess.   They would appear to be degenerate parasitic forms, having cut everything down to essential characteristic nucleo-protein, rather than primitive forms evolved independently of higher organisms.

Their significance here is that they illustrate the effects on physical structure of the existence of large regular molecules.   Once the size of 100 Ångströms or so is surpassed, new kinds of interactions, imperceptible against the thermal background at smaller sizes, become apparent.   These are the long range forces assumed to account for many colloidal phenomena, such as gel formation and coacervation.   Tobacco mosaic virus provided the first quantitative pictures of such forces, because it showed that the identical rod-shaped particles of the virus maintained themselves at equilibrium distances which are dependent on the pH and salt concentration.   The physical nature of these forces, though almost certainly connected with ionic atmospheres, is still in dispute.   Langmuir, Levine, and Verwey and Overbeek (1948) have produced theories which account for them qualitatively, but of their real existence there can be no doubt.

This statement of course does not exclude the possibility that no physical forces exist, but instead that thermodynamic equilibria are satisfied for particles maintaining certain distances from each other. What is important for biology is that such distances can be maintained in media such as those of cells.

The moment, therefore, that macro-molecules of this type are produced they must interact in the liquid and give rise to the physical unity which characterizes the individual living organism. Because of the vagueness of the terms, it is difficult to say whether we should count the origin of life from the first moment that steady interactions occurred between complex molecules in a general medium, or from the point when a section of this medium was separated out, sufficiently large to contain a self-maintaining system of reactions with the medium. We might, if we wish to be precise, qualify the first with the origin of *life*, the second with the origin of *organisms* or living things. The first stage is difficult for us to grasp. The idea of life without living things may seem a contradiction in terms, and yet the evidence, as far as it goes, indicates that such a stage must have occurred because simple molecules must have preceded complex ones, and because definite organisms cannot be formed without the pre-existence of complex molecules*.

For the purposes of understanding the early stages of development of life it is, however, not absolutely necessary to stipulate the colloidal state of the system. For one thing that is becoming increasingly clear of recent years is that the chemical reactions and the modifications of these reactions are the most characteristic features of vital activity. We must therefore infer a long stage of chemical evolution, of what might be called the internal economy of life, long before we need consider the external shape or forms of function of living things. There is work now to be done on the chemical evolution of life more difficult but also far more fundamental and useful than that which Darwin did on the evolution of the higher forms. The key to the understanding of this chemical evolution of life lies in the junction between observational biochemistry on the one hand and quantum theory on the other.

The basic chemical problem of all vital transformations is the achievement of chemical changes in an isothermal medium, by which large amounts of energy can be made use of in the small steps which alone are permissible in such a system. It is found, for instance, that in the linked enzyme systems which occur in all vital chemical transformations, the individual steps are never more than some sixteen calories and are often much less, down to three calories, hardly more than that of a hydrogen bond, whereas the total exchanges of energy may be several hundred calories, equivalent to the complete combustion which, without these enzyme mechanisms, could only occur at high temperatures. In providing these small quantum jumps, the existence of particularly labile inorganic reactions is obviously of the most critical importance. Two of these, on account of their extreme abundance, are known to be specially important—the oxidation and reduction of ferrous to ferric iron, and of sulphydsyl, —SH, to disulphide, —S—S—. Early life, in the absence of atmospheric oxygen, must have proceeded almost entirely by the enzymic utilization of these transformations.

The processes of life have a close analogy to those of a chemical factory, only here, instead of the materials being poured from one reaction vessel to another, the individual molecules diffuse from one enzyme to the next, the rates are fixed,

---

* I owe this example of such a system to Haldane. The bacterium *Haemophilus caris* cannot synthesize hematin and *H. influenzae* cannot synthesize coenzymes I and II. Neither can grow alone on peptone, but a mixed culture grows well.

and a number of them circulate in these cycles, making use of a certain fraction of the available energy of the reaction, to reverse the entropy gain. As Schroedinger has said, life consumes not food, but negative entropy, and it is able to do so by the existence of what are effectively solid structures in the protein molecules themselves. It is with the establishment of any one such chain of reactions where complex molecules can be fed in at one end and simple ones liberated at the other, with a net energy gain, that we may imagine that living processes started.

We have seen that the complex molecules themselves may have been produced by straight light absorption, but on the other hand the environment did not contain any large sinks of energy such as are provided by the existence of free oxygen molecules in the present atmosphere. The first life processes, therefore, must have been extremely inefficient, similar to those that exist at the moment in anaerobic fermentation, where only 10 per cent of the available free energy is made use of. Nevertheless, once a process of this sort gets started, it is bound to spread. We know at the present day a number of bacteria which are called autotrophic, that is, which can nourish themselves entirely on inorganic ions, nitrate, phosphate, sulphydryl etc. These, however, cannot be representative of the primitive organisms, because they contain all the most complex enzymic systems.

It is certain that life in its first stages was entirely dependent on the pre-existence of organic molecules. But, correspondingly, as life spread, these organic molecules must have been consumed, and if no other process had supervened, early life must have burnt itself out. This, indeed, may have happened over and over again, because such burning, i.e. the decomposition of all complex molecules, must have restored the original state of affairs and started the whole process all over again, without leaving any notable traces. However, we know by our own existence that they did not all burn themselves out.

It seems most probable that the first crisis of primitive life was resolved by some organism chancing on another mode of collecting energy for chemical reactions, namely one which made use of the current input of solar light. At what stage this happened depends on whether the first organisms were adapted to work in the light or the dark. The latter seems more probable, especially as most of the reactions that we now know of in biochemistry are stopped by short-wave ultra-violet light. It is still, however, quite possible, either in the sea or, more probably, in a mud-bank, for the surface to be exposed to the full flux of sunlight, while the interior is carrying on with much reduced light intensity or in virtual darkness. Life creeping up towards the light might in this case find itself involved in the processes in which the absorption of medium ultra-violet or even visible light might assist the cyclic transformations and make them capable of utilizing less highly formed molecules. This is apparently what happens in the purple sulphur bacteria of today, which utilize simultaneously the energy of absorbed light and that derived from oxidations and reductions of iron and sulphur compounds.

Effectively such a process is likely to occur the moment coloured substances capable of acting as sensitizers, as in photography, are synthesized. One group of these which seems to be especially favoured, possibly on account of their easy production, are the *porphyrins*, formed by condensing four pyridene derivatives around a metal ion. It is extremely suggestive that these porphyrins are found acting simultaneously as light receivers in the chlorophyll of all plants, and as respiratory pigments in cells, as oxygen carriers in many animals, and finally,

and most strangely, associated with nitrogen absorption in the root nodules of peas. This points to all life being derived from a light-absorbing intermediate form—essentially a plant.

From the light-assisted metabolism of the purple bacteria to the full photosynthesis may not be a very wide step, but once the latter process is achieved, the future of life is assured. Modern research on synthesis shows that it is far from the simple reaction $CO_2 + H_2O = CH_2O + O_2$; that in fact the three processes of light absorption, carbon dioxide absorption and the liberation of free oxygen occur quite separately and may not all have been evolved at the same time. However, once photosynthesis started it made its mark on all subsequent living things, and was undoubtedly so successful that it wiped out earlier forms of life, except in very obscure holes and corners. One characteristic product of photosynthesis is sugar, and the universal prevalence of suger is only less marked than that of protein in the organic world, though all indications point to it as a much later derivative. Combined either with phosphorus or with purine bodies, or with both, sugars take part in an enormous number of enzyme systems, and particularly in nucleic acid, with its close associations with growth and reproduction.

The processes of respiration are so closely attached to those of photosynthesis that it seems probable that both occurred together in the first place, and only later became separated, when organisms developed predominantly one or the other process. Respiration, as already indicated, is a very much more efficient process than fermentation, and immediately increases the activity and range of living things. But full-scale respiration cannot have occurred until photosynthesis had been going on for some time and had produced the necessary oxygen. Once this happened there were two other direct consequences. In the first place, as already indicated, the oxygen drifting into the upper atmosphere interacted with the ultra-violet to produce an ozone layer which effectively blocked all the original syntheses on which life existed. But the very nature of photosynthesis prevented this having any serious effect on the production of organic from inorganic material. The danger may well have been in the other direction, a relatively unchecked building-up of carbon into cellulose and other sugar products. But the simultaneous existence of oxygen and combustible hydrocarbons put a premium on mechanisms of the reverse kind, and organisms specializing in respiration : bacteria, fungi and animals, tended to restore the balance. This is indicated in the diagram by a second maximum in the effective utilization of sunlight brought about by the reversal of the upsurge due to photosynthesis, by the increase of catabolic processes leading ultimately to the equilibrium in which we now live.

From the moment of the existence of an oxygen-rich atmosphere, the physical environmental conditions of life became essentially what they are now. This is as far as it is possible to go, at least at present, on the basis of purely physical and chemical knowledge, even assisted by what we know of biochemistry. For understanding development in the third act, that of organized life, we must rely on the ever deeper study of living organisms, both in structure and function, and use physical and chemical knowledge to interpret what is observed. A hundred years ago the discovery that the cell was the unit of life seemed to explain the common origin of the large observable organisms and tissues which were then studied in terms of some biological atom. In one sense this is still true. With the exception of the viruses, and possibly of some bacteria and the mycetozoa, all animals and plants are composed of cells ; but the structure of the cell is likely to

prove as complicated in terms of its constituents as that of the organism in terms of cells. The first fact which makes a science of cytology possible is that of cells reconstructed on the same basic pattern and developed, and particularly reproduced, by a set of complicated phenomena identical over the whole living kingdom.

The evolution of organisms must have been preceded by a period of the evolution of the cell. But the story of that evolution is certainly the most difficult part of the whole unravelling of biological processes, because it is at the same time the furthest removed from synthetic chemistry and analytical biology. Nevertheless in recent years, thanks to the new methods described above, much deeper understanding of the cell is becoming possible. The most striking general feature is the dual construction. The cell consists, inside a membranous envelope, of two parts, a nucleus itself provided with a membrane, and an external cytoplasm.
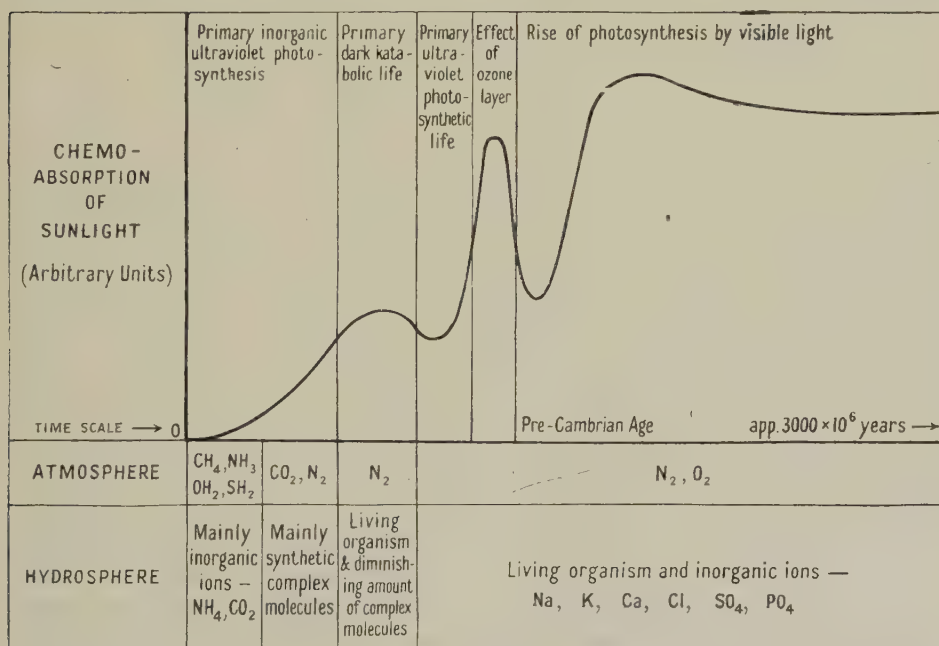
This figure is entirely hypothetical except as to the sequence of the stages. The relative heights of the maxima may have been different and the stages may have overlapped much more than is shown, even to the extent of blotting out the maxima and minima.

Each of these volumes in turn contain other finer parts. The nucleus contains filamentous processes of varying number, the chromosomes, together with the associated nucleoli and centromeres, while the cytoplasm contains a world of particles of varying dimensions, particularly the mitochondria, the golgi bodies and a number of the smaller plastids, including the chloroplasts in the plants and the microplasts and other so-called organelles.

Many, if not all, these constituents appear to be self-reproducing. Therefore even in single-celled organisms there is considerable differentiation. Protoplasm is therefore far from a structureless substance, and to its microscopic complexity must correspond an even greater diversity of chemical and physico-chemical function, all pointing to an equally complicated history. I will not attempt to discuss the former of these, the cell equilibrium brought about through enzymes

which seem to be attached to the mitochondria, and to the apparently dominating rôle of desoxyribose and ribose nucleic acids, the former being synthesized in the nucleus and its amount determining growth and division.

The cell, chemically speaking, may be considered as a little world, if not two little worlds, in which the whole set of chemical reactions characterizing some large volume of pre-organic life is concentrated. The essential feature of the chemical life of the cell is its extreme persistence along established lines of chemical reaction, and the way in which these characteristic reactions are reproduced over and over again in daughter cells. Schroedinger has pointed out that this exact reproduction is a quantum phenomenon in itself, and has attached the possibility of over-coming the entropy-increasing effect of normal diffusion by the quasi-solid (solid at least in one dimension) structures of the chromosomes themselves. Actually this uniformity of chemical action is by no means absolute, particularly in bacteria. The actual biochemistry of the cell can in fact be very violently varied by changes in the environment, and even the nucleus itself and the genetic processes are, as we now know, capable of random modification and of directed modification by specific chemicals. Their isolation in the past depended largely on the difficulty of getting the chemicals to the places where they could act.

Where physical methods can be of use is in interpreting the structures visible in the cell and in inferring others. Protoplasm is by no means a uniform substance. Electron microscope and x-ray researches show that it contains, quite apart from the relatively large structures previously referred to, others of intermediate size which, on account of their greater surface, are probably the seat of the greatest activities. The viscosity of the cell contents shows that it contains an appreciable amount of fibrous molecules. These are almost certainly of protein nature. One of the most interesting discoveries in this field is the extreme ability of the so-called globular, but more probably plate-like, proteins to aggregate in the form of fibres. This aggregation, if carried out without drastic conditions of temperature or acid, seems to be perfectly reversible. At least it has been observed to be so · in the cases of insulin, actin, tropomyosin and fibrin. It would appear indeed that practically all fibres found in biological systems are made in this way from particulate proteins, for example paramyosin, where the particles can be seen in the electron microscope, collagen and keratin. Conversion of particulate to fibrous proteins is probably some form of chain-ring polymerization, but in the more labile forms this can only affect a small number of the available bonds. These fine fibres, running from 100 to 200 Ångströms in diameter, have of course very large surface areas and are capable, therefore, of carrying on almost as active chemical interchanges as free molecules. Indeed, it seems very probable that all the enzymes in cells are attached in this way, the most notable being the adenosin triphosphatase, which seems to be bound by myosin and to play an important part in muscle contraction.

An analogous phenomenon follows necessarily from the very existence of such fibrous molecules. Zocher and Birnstein (1929) first showed that in dilute clay suspensions the particles tend to arrange themselves in parallel sheets at regular intervals, which may be as great as 5000 a. I was able to study some phenomena in tobacco mosaic virus, as mentioned above, and in fact we can say now that all such fibrous molecules are attached to each other by long range forces, the range of which is dependent on the conditions of the medium, which may impose coagulation into microscopically visible fibres when the distances are small to a practically fluid and orientated gel when they are large.

The variability of the degree of fibrous nature of the cell contents is nowhere better shown than in the phenomenon of cell division, a complicated pattern which, however, is substantially the same for all cellular organisms. Some ten years ago I put forward a theory of mitosis based on the observation of these long range forces, and particularly on the analysis of the structure of the spindle-shaped tactoids (Bernal 1940). These were first observed by Zocher in iron oxide sols and later studied in more detail in tobacco mosaic virus. When a uniformly diffused area of rod-shaped particles is caused to aggregate, the forms taken by the aggregation are no longer spherical drops, but, on account of the long range forces, particles tend to remain in parallel orientation, giving rise to anisotropic surface forces and leading to an equilibrium in the form of a tactoid with a geometrical surface obeying the formula $\sigma_1/r_1 + \sigma_2/r_2 = \text{const.}$, where σ_1 and σ_2 are the principal surface tensions and r_1 and r_2 the principal radii of curvature. Exactly the same form is taken by inclusions of clear solution in a concentrated aggregate of long molecules. These, analogous to bubbles, are called negative tactoids. Such tactoids undoubtedly exist in dividing cells (see Bernal and Fankuchen 1941).

It would appear that the mechanism of their formation is somewhat as follows: at the instant of division, spontaneously or, perhaps, stimulated by changes going on in the nucleus, a small particle in the cytoplasm, the centrosome, causes a ring-chain polymerization of the proteins, proceeding as in insulin and forming a spherulite assembly of fibres radiating from a point. The centrosome divides and the spherulite correspondingly becomes a tactoid, or spindle. That this is the case has been shown very beautifully by the photographs of mitosis in polarized light. The existence of the tactoid produces an effective long range forces field. In such a field particles which do not mix with the system, such as the chromosomes, will tend to be driven to the equator where the particles are most parallel. This is the stage known as the metaphase.

The next stage marks the reversal of this process, but this appears to originate in a special section of each chromosome, the centromere. The centromeres of each paired chromosome, for the chromosomal division occurs at an earlier stage, proceed to separate, but in such a way as strongly suggests that each pair are the poles of a negative tactoid, and are in fact centres of disaggregation of the protein fibres. Thus they, so to speak, eat their way away from the equator towards the poles. That considerable forces are involved is shown by the fact that anomalous chromosomes which have not fully divided are literally torn in two. Much more research is needed to show whether this hypothesis is the correct one, but it has at least the merit of not invoking any forces other than those that can be demonstrated in other organic and even in inorganic systems such as ferric hydroxide and vanadium pentoxide gels.

Similar explanations based on the knowledge of long range forces may go far to explain the mechanics of chromosome pairing itself. In reproduction the chromosomes belonging to the father and mother nuclei are together in identical pairs. These must, in fact, seek each other out and arrange themselves mutually with extraordinary accuracy. The variation of nucleic acid material along the chromosomes, with its complicated pattern and the corresponding variation in long range forces, may serve to explain this process. Full interaction can only occur if there is a proper opposition of parts, and the difference between a good fit and a bad may insure that after a number of trials a fit is achieved in one place and then proceeds by zipper action over the rest of the chromosomes. We are only

at the beginning of these studies, and far greater complexities will probably be revealed by further observation.   But these complexities will act as a spur to further physical researches leading to their explanation.

So far I have spoken only of the interior of the cell.   But for the survival of evolution of life, the relation of the cell to its environment seems now no such great mystery as it seemed to Henderson, considering, if our view is correct, that it was blocked out of that environment and represents simply a selective enclave of balanced chemical processes.   The pre-condition of the existence of a cell is that it can maintain the concentration of its elements in a medium of very much lower concentration.   This involves work, but it is a kind of work that the protein enzyme systems in the cells are particularly suited for, and for which the properties of lipoid membranes provide the necessary physical scaffolding.

In the further evolution of organisms we can distinguish size as the major conditioning factor.   The uni-cellular organisms of a few microns in dimension must necessarily live in a world in which molecular forces rather than hydro-dynamic ones are of prime importance.   Small molecules on which nutrition and respiration depend must enter freely by diffusion.   But diffusion is absolutely limited by size, and the larger an organism is, the smaller its effective surface, and the smaller the effective volume from which it can draw its nourishment.   There are two ways out of this dilemma, and both have been followed by different organisms. The first is to grow only in one or two dimensions, to produce filamentous or leafy forms keeping a large surface.   The second is motility, not the directed motility of the larger organisms, but simply moving for moving's sake, out of the comparative starvation of the immediate neighbourhood.   The ceaseless motion of bacteria and protozoa had originally no other point, but once a controllable motile mechanism is evolved from such simple protein molecules as cilia, hardly more than a bundle of some eight or sixteen macro-molecules, future evolution towards directed motion becomes almost a necessity.

At this stage there are two directions open to organisms.   One is to improve their biochemical mechanisms and be able to make use of a very dilute environment, and the other to move into a better environment ; that is the basic distinction between plants and animals.   The first remain absorbers of individual molecules, and their motility is limited to extending their surfaces and arranging for the slow and steady flow of an internal environment.   The motile forms, however, sooner or later are bound to bump into each other, and from that contact arises the possi-bility of digestion and, therefore, the possibility of acquiring new material in prefabricated pieces rather than molecularly.   Motility in itself may fluctuate between total motility, in which the organism moves through the environment, and relative motility, in which the organism stays still and pushes the environment past it, as in the ciliates, sponges, oysters or barnacles.   In the first case the essential new feature is direction, a movement along a chemical and thermal gradient.   This involves the production of some kind of sensitive element at the head end, and some link between this and the motility.   From these simple begin-nings follows the evolution of the necessarily linked triad—sense organ, nerve system and effector.

The actual evolution of the higher organisms, using cell division to create a larger organism than is possible with single cells, may have been really only a mechanical necessity.   Once the very small dimensions of the single cell are surpassed, forces utterly inappreciable on the micro-scale become important, such

as turbulence of the medium and gravity. To resist these, greater strength than is provided by the protoplasmic framework is needed, and this is available by using the mechanical strength inherent in cell walls, either simply or strengthened by protein fibres or by cellulose. Where additional strength is required, mineral deposits like shell and bone can be added. Multi-cell organisms can afford to specialize, and our understanding of the physical nature of these specializations, such as the passing of nerve currents, the contraction of muscles or the secretion of chemicals, seem to show that they all use rather specially selected aspects of mechanisms common to all protein systems.

But now the particular function of the organ is no longer self-determining and absolute. It is part of a larger coordination. Modern mathematical analysis is showing us that the very fact of coordination, the very linkage mechanism itself, introduces new complexities and new possibilities of a very specific kind. Take, for example, one of the most complicated, the mechanism of vision. It is relatively straightforward to arrange a one-to-one connection between a single receptor in the retina and a single effector in the visual cortex. But this is only the beginning of the story. Such a system could see, but it could neither recognize nor follow. Three layers of cells are involved in grouping and sorting out the impulses of the cortex and creating the percepts of objects which are the basis of sensual experience of the higher animals. The arrangements of these layers of cells, as Weiner has shown, are essentially similar to the arrangements in modern electronic calculating machines with similar functions. An even higher complexity is required if we have to move from percept to concept, a process which in its fullness has only occurred in man, or possibly in some other social animals. Here the mutual relations of organisms originating in reproductive links have led to a still higher order of mutual interaction and corresponding permanency of cultural tradition.

This brief survey of the field of life does show that we are now beginning to see the possibility of accounting for it in terms that fit in with our intellectual and manipulative control of the non-living parts of environment. It is certain that this account will appear as absurd to those who come after us as did those of Borelli or of the early corpuscular mechanists of the seventeenth century. Yet even then this approach to rationality did make possible further advances of biology at that time, and so we may hope from the present analysis to achieve corresponding advances in biophysics. This does not mean any reduction of biology to simple physics. The complexity of the successive evolutionary levels mark definite orders of advance each with its own laws and each with its own internal unbalance leading forward to new levels. However, by elucidating structures, by following out mechanisms, we may at least clear the air of mystical interpretations which serve simply to conceal ignorance. Instead of concealing ignorance we may hope to reveal it, to find the places where the explanations will not fit, and sooner or later to find the means for studying just these points. In such a study we may reasonably expect to discover further and further points of ignorance and difficulty. But even if the task were endless it would still be worth while. Every new relation gained is an achievement of order, giving confidence and direction to future work, and at the same time providing for new controls over our biological environment, including our own bodies. The profundity of physics should not blind us to its relative simplicity in the scheme of human knowledge. Here in biophysics we have something so complex that it will require not only more industry but more intelligence than went into the

earlier stages either of biology or of physics. Man has, in the past, used his knowledge to effect gross changes in his environment, first mechanical and then chemical. Only an understanding of the biological and social sciences can show how to make those changes of real benefit to man, and how to balance power with intelligence.

REFERENCES

ASTBURY, W. T., 1933, *Fundamentals of Fibre Structure* (Oxford : University Press).
BERNAL, J. D., 1940, *The Cell and Protoplasm*, ed. F. R. Moulton, (Washington : Amer. Assoc. Adv. Sci.), p. 199.
BERNAL, J. D., BAWDEN, F. C., PIRIE, N. W., and FANKUCHEN, I., 1936, *Nature, Lond.*, **138,** 1051.
BERNAL, J. D., and FANKUCHEN, I., 1941, *J. Gen. Physiol.*, **25,** 111.
CROWFOOT, D., 1941, *Chem. Rev.*, **28,** 215.
DARWIN, F., ed., 1887, *The Life and Letters of Charles Darwin*, vol. 3, p. 18.
DAUVILLIER, A., 1947, *Genèse nature et évolution des planètes* (Paris : Hermann), pp. 169 ff.
ENGELS, F., 1940, *Dialectics of Nature* (London : Lawrence and Wishart), p. 194 ff.
GOLDSCHMIDT, V. M., 1923–1937, *Skriften Norske Vidensk. Akad. K.*, **1.**
HALDANE, J. B. S., 1928, *Rationalist Annual*, " The Origin of Life ", pp. 148–153.
HAST, N., 1947, *Nature, Lond.*, **159,** 354.
HUXLEY, T. H., 1901, *Collected Essays*, " On the Physical Basis of Life ", vol. 1 (London : Macmillan).
KUHN, W., 1946, *Experientia*, **2, 10.**
LWOFF, A., 1943, *L'Évolution Physiologique* (Paris : Hermann).
MACEWAN, D. M. C., 1948, *Trans. Faraday Soc.*, **306,** 349.
OPARIN, A. I., 1938, *The Origin of Life* (New York : Macmillan).
PERUTZ, M. F., 1949, *Proc. Roy. Soc. A*, **195,** 474.
PIRIE, N. W., 1937, *Perspectives in Biochemistry*, eds. J. Needham, and D. R. Green, *The Meaninglessness of the Terms 'Life' and 'Living'* (Cambridge : University Press), pp. 21 ff.
PRIGOGINE, I., 1947, *Étude thermodynamique des phénomènes irréversibles* (Paris : Dunod).
TURKEVICH, J., and SMITH, R. K., 1946, *Nature, Lond.*, **157,** 874.
VERWEY, E. J. W., and OVERBEEK, J. Th. G., 1948, *Theory of the Stability of Lyophobic Colloids* (Leyden : Elsevier).
ZOCHER, H., and BIRNSTEIN, V., 1929, *Z. Phys. Chem. A*, **142,** 113.

The Contributions of Thomas Young to Geometrical Optics, and their Application to Present-day Questions

By T. SMITH

The Thomas Young Oration for 1949, *delivered* 10th *June* 1949

ABSTRACT. Thomas Young's constructions for primary and secondary conjugate foci are described, and his algebraic equations are given. The latter are put into the form found most convenient when an instrument of several surfaces is considered. The special forms of these equations referred to the centres of curvature of the reflecting or refracting surfaces are used to show that in spherically symmetrical instruments the number of degrees of freedom—which is equal to the number of surfaces when these are spherical—must exceed the number of aberrations that can be corrected. It is concluded that non-spherical surfaces are likely to be very important or perhaps essential for any marked improvement in the correction of optical instruments, and that Young's equations, which relate expressly to spherical surfaces, may prove less useful in the future than they have been in the past.

THOMAS YOUNG, in whose honour this lecture is named, was unquestionably one of the most distinguished men of his time. Our knowledge of the man himself, who was by no means without peculiarities, depends largely on the biography written by George Peacock, Dean of Ely and also Professor of Astronomy at Cambridge. Like others of his age, Young's work covered a very wide range, much of it outside the field of science. His most important scientific publications—three Bakerian Lectures and two courses of Lectures on Natural Philosophy at the Royal Institution—appeared from 140 to 150 years ago. Some of his views, particularly those on the nature of light, were violently assailed at the time because they conflicted at points with those formed by Newton; and the Royal Society itself was strongly criticized for publishing such perverse doctrine. On these and some other matters Peacock has much to say, but Young's contributions to geometrical optics were evidently regarded as of little importance, for they are not even mentioned. Perhaps this is to be expected, since Young, as the inscription in Westminster Abbey reminds us, was 'a man alike eminent in almost every department of human learning'. But I think these contributions have proved of sufficient importance for us to recall them when we are reminded of Thomas Young's own work and of what has followed from it.

Our textbooks of physics have little to say about Young's contributions to our knowledge of optics. In geometrical optics students probably know Young's construction for finding the refracted ray corresponding to any ray incident on a spherical surface, though it is sometimes ascribed to a different discoverer. The construction is as follows (Figure 1): Let C be the centre of curvature of a spherical refracting surface of radius r which bounds two media of refractive indices μ and μ'. Suppose a ray lying in the former medium intersects the refracting surface in S and the sphere of radius $r\mu'/\mu$ centred at C in A. The refracted ray is SA' where A' is the point of intersection of the straight line CA with the sphere of radius $r\mu/\mu'$ having C as its centre. It may seem surprising to us that a construction so simple and convenient was not known before Young's time; we have to remember that we are almost at the beginning of the theory of optical instruments as we know it today.

This construction is the foundation for developments of great importance. If M is any point on the incident ray through S, the straight line MC will intersect the refracted ray in some point M′; and in virtue of the complete symmetry of a sphere it is clear that all rays from M which make an angle with MC equal to that made by MS will also pass through M′ after refraction. In conformity with the usual phraseology of geometrical optics M′ is the image of M so far as rays such as MS are concerned. It is not in fact the only image. To distinguish it from others it is called the secondary or sagittal focus. The title must not lead us to think that it is of less importance than some other focus—in reality the reverse is the case in many instruments. The point to be borne in mind is

$$SP = u, \quad SP' = u'$$

$$x'y'p'$$

$$a = \frac{\mu' \cos \phi' - \mu \cos \phi}{r}$$

$$\frac{\mu}{u} + \frac{\mu'}{u'} = a$$

$$\frac{\mu \cos^2 \phi}{v} + \frac{\mu' \cos^2 \phi'}{v'} = a$$
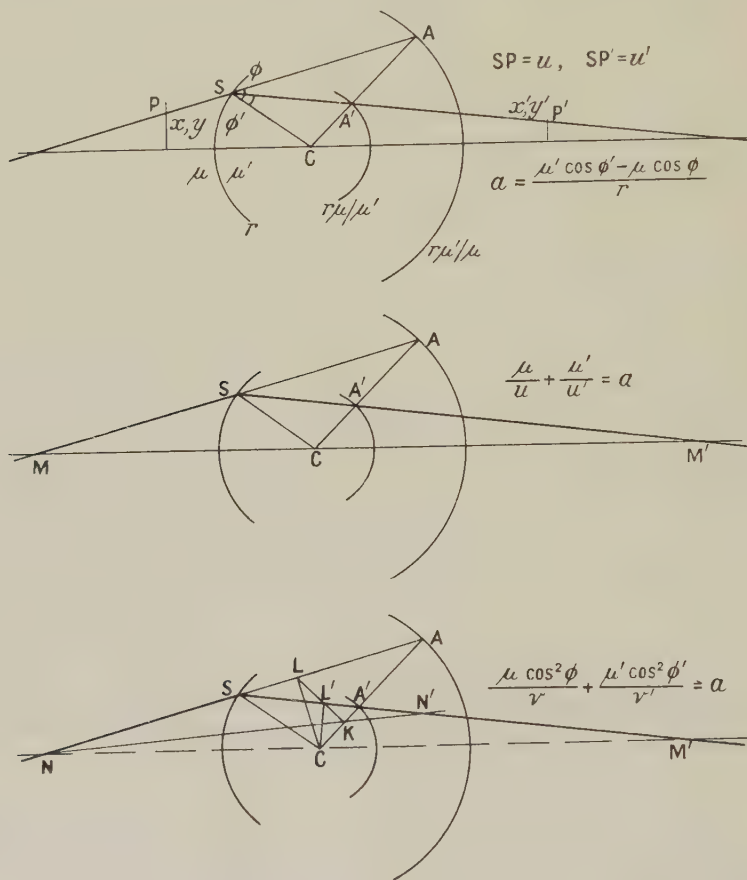
Figure 1.

that C is a centre of projection for the determination of conjugate secondary foci.

This construction does not hold for the points of intersection of rays which lie in a plane passing through the centre of curvature C, but Young obtained a construction for this case also. He showed that corresponding points could be obtained by projection through another point K. It is easy to show that if L and L′ are the feet of the perpendiculars from C to an incident and refracted ray passing through the point S of the refracting surface, an incident ray through L which is refracted at a point near S in the same plane of refraction will subsequently pass through L′. The points L and L′ are therefore conjugate to one another for a

narrow pencil of rays lying in this plane—commonly called the primary or tangential section. The points A and A′ on the concentric spheres already mentioned are conjugate for all rays, and S is a self-conjugate point. Since three pairs of points determine the conjugate relation completely, the primary conjugate foci can be found by projection through K, the point in which LL′ intersects CA. Thus in Figure 1 N′ is the primary image of N on the ray NS, where N′ is the point of intersection of NK and the refracted ray. The figure shows that the distance between N′ and M′, the primary and secondary images of N, may be large if the angle of incidence at S is not small. The distance CK itself is equal to $r \sin \phi \sin \phi'$, where ϕ and ϕ' are the angles of incidence and refraction; for small values of ϕ this separation is almost negligible, and the want of sharpness in an image is correspondingly small; but when incidence is decidedly oblique the aberration in an image tends to be large.

By these graphical constructions, it is possible, with careful drawing, to obtain a very fair idea of the aberrations present in an optical instrument. Curvature and astigmatism in particular are determined directly. The importance of this procedure is largely due to the fact that C and K are the limiting points of the locus of the centres of projection for all rays in the neighbourhood of the principal ray considered.

But Young was not content to leave the subject at this stage. However informative graphical methods may be, they are inadequate when great accuracy is needed. This requirement of high accuracy is fully met by Young's algebraic representation of the relations he established. If u, u' denote the distances SM, SM′, and v, v' the distances SN, SN′ as shown in the figure, he derived the equations:

Secondary relation
$$\frac{\mu}{u} + \frac{\mu'}{u'} = a$$

Primary relation
$$\frac{\mu \cos^2 \phi}{v} + \frac{\mu' \cos^2 \phi'}{v'} = a,$$

where
$$a = \frac{\mu' \cos \phi' - \mu \cos \phi}{r}.$$

The importance of these equations may be realized from the fact that they are still in general use, in the form given by Young, by lens designers to determine the positions of the focal lines of an image. Any consistent convention for the signs of u, u', v, v' may be employed—the one adopted here, which seems to me the simplest, is to use a positive value for distances measured from the surface into the medium itself. This is equivalent to the convention that a positive sign corresponds to a real immediate object or image, and a negative sign to a virtual immediate object or image. It should be particularly noted that no frame of coordinates is employed. The equations involve only significant geometrical quantities, and all lengths are measured along the rays themselves. The direction in which the light travels is also not significant. When light is reflected instead of refracted we have only to note that $\phi' = \pi - \phi$. This is not the same thing as changing the sign of μ'.

To the present-day physicist these contributions of Young's, according to the bare outline I have given, may not appear of much account. But we must keep in mind the state of knowledge in his time. Of the common aberrations, spherical aberration and longitudinal chromatic aberration were well enough

known; but the existence of the distinct error coma was not suspected and the word itself had not been coined; the fault we know as astigmatism had been noted by Newton, but precise knowledge of it is due to Young. It is significant that on the Continent, where his work remained unnoticed, the properties of the point K as a projective centre were only discovered some sixty years later. According to von Rohr the theory of astigmatism, as given by Coddington on the basis of the researches of Young and Airy, was widely known in this country long before it attracted any attention abroad. I am not aware that our instrument makers made use of this knowledge to achieve a standard of performance surpassing that of their continental contemporaries.

I have mentioned that Young's equations are still used by lens designers, and might add that I understand that work is now in progress on the extension of these or similar relations with the aim of making them applicable to pencils of greater size than those for which Young's equations are valid. It is hoped that these developments will lead to the construction of photographic lenses of large aperture which will give fine definition over a larger field of view than has yet proved possible.

Young's equations necessarily relate to a single reflecting or refracting surface, but optical instruments in general have many surfaces. The procedure usually adopted is to start with an object in a known position, so that the u and v for the first refraction are known. From the equations the values of u' and v' are found which determine the positions of the primary and secondary images. These points are then regarded as the object points for the second surface, the sums $u_1' + u_2$ and $v_1' + v_2$ being equal, to the distance between the first and second surfaces measured along the ray. This process is repeated at each surface in succession, until the final u' and v' are found, giving the positions of the foci for the complete instrument. The information derived in this way refers to one definite object point. To obtain results as comprehensive as those given by Young's single-surface formulae it would be necessary to carry through this set of operations for not less than three suitably chosen object positions. A better alternative is to put the equations in another form which involves less labour and at the same time gives additional information.

In the new equations we avoid finding the distances between the surfaces and the images formed between the first and last surfaces of the instrument. The only quantities we require to know are the powers $a_1, a_2, \ldots a_n$ of the several surfaces and the distances between successive surfaces measured along the ray. The magnitude it is most convenient to use is this intercepted length divided by the refractive index of the medium between the two surfaces; this is invariably given a negative sign if the usual convention is adopted by which the power a is given a positive sign when the surface tends to converge rays incident upon it. The letter d is used for this negative product, and the addition of the suffix k implies that d relates to the separation of the kth and $(k+1)$th surfaces. By avoiding the longitudinal distances u and v, which may vary greatly in magnitude, and utilizing in their place the distances between surfaces, which are all of moderate magnitude, we obviate a possible source of numerical inaccuracy.

The new equation for the secondary fan of rays differs from Young's in that use is made of cartesian coordinates. We assume that the instrument considered has an axis of rotational symmetry, and we choose this as the z axis of an orthogonal system of cartesian coordinates x, y, z. We require directional variables as well

as point coordinates, and take for these the products of the direction cosines of a ray and the refractive index of the medium concerned; denote these by ξ, η, ζ. In place of Young's secondary equation we then have the new equation

$$\begin{pmatrix} \xi' & -x' \\ \eta' & -y' \end{pmatrix} = \begin{pmatrix} \xi & -x \\ \eta & -y \end{pmatrix} \mathscr{S}$$

for a system of any number of refracting surfaces, where $\mathscr{S}$ is a matrix of the form $\begin{pmatrix} b & d \\ a & c \end{pmatrix}$, and is defined by the product

$$\mathscr{S} = D_0 S_1 D_1 S_2 \ldots . S_n D_n,$$

where D_0 represents the displacement from the object point (x, y) to the first surface, S_1 the first refraction, D_1 the displacement from the first to the second surface, and so on, and D_n the displacement from the last surface to the point (x', y') on the emergent ray. The definitions of the factors are

$$D_k = \begin{pmatrix} 1 & d_k \\ . & 1 \end{pmatrix}, \qquad S_k = \begin{pmatrix} 1 & . \\ a_k & 1 \end{pmatrix}.$$

In this notation the point (x', y') will be the image of the object point (x, y) if x', y' depend on x, y but are independent of ξ, η. That is to say the condition that the terminal points for which $\mathscr{S}$ has been computed are conjugate to one another is that the element d of $\mathscr{S}$ is zero. It is readily seen that for a single surface the condition $d = 0$ is identical with Young's secondary equation. Obviously the new equation gives much more information than Young's, and its use is made very elastic by regarding D_0 and D_n as variable factors to be associated with the computed product $S_1 \ldots . S_n$.

It will be noted that though we introduced longitudinal as well as lateral coordinates in the first place the quantities z and z' do not appear in the equation. There are definite connections between the transverse variables, and distances measured along the rays themselves are also important, but measurements along the axis as such have no optical significance. There is accordingly no virtue in imposing conventions relating to object or image distances measured along the axis; but it is of the greatest importance, as Young's own equations show, to define clearly what signs are to be given to distances such as u, u', v, v' measured along a ray. I emphasize this distinction because it seems to be largely neglected by teachers today.

Young's primary equation may be treated in a similar way. This relates essentially to a near ray lying in the same axial plane as the principal ray. If the inclination of the near ray to the principal ray in the object space is θ and in the image space is θ', and the distances between the rays at the end points determined by D_0 and D_n are p and p' respectively, the new primary equation is

$$(\mu'\theta' \quad -p') = (\mu\theta \quad -p)\mathscr{T}$$

where
$$\mathscr{T} = D_0 T_1 D_1 T_2 \ldots . T_n D_n$$

and
$$T_k = \begin{bmatrix} \dfrac{\cos\phi_k}{\cos\phi_k'} & . \\ \dfrac{a_k}{\cos\phi_k \cos\phi_k'} & \dfrac{\cos\phi_k'}{\cos\phi_k} \end{bmatrix}.$$

As before, Young's equation corresponds to the vanishing of the d element of $\mathcal{T}$. The computation of the $\mathcal{S}$ and $\mathcal{T}$ matrices is of special value in the design of optical instruments. The $\mathcal{S}$ matrix in particular can be employed in tracing rays through surfaces of any shape—a development not contemplated by Young.

I wish now to turn to some questions concerning optical instruments which are of special interest to us today, and to consider them with particular reference to the equations relating to primary and secondary foci. Most instrumental problems are far too involved and difficult to be discussed on such an occasion as this. Aberrations are far from few, and lens designers spend their lives battling with them—often, I think they would admit, with somewhat indifferent success. Instead of attacking so formidable a problem, we must contrive to evade it. This will be possible if we can base our inquiry on instruments which from their very nature are free from all but a few aberrations. Happily there are such instruments, and even if the problems they present appear somewhat trivial to the professional designer they are nevertheless of interest and will serve our purpose.

The questions I propose to consider are the relation between the number of constructional degrees of freedom in an instrument and the number of aberrations that can be corrected; and the importance for future instruments of non-spherical refracting or reflecting surfaces.

In the last few years much interest has been taken in telescope objectives designed to have a spherical image which is as nearly as possible equally well corrected at all points over a considerable angular field. This equality is achieved by a construction which aims at symmetry around a centre. If complete spherical symmetry were possible it is obvious that the elimination of aberrations in the image of any one distant point would imply correction of the whole field. In known instruments this ideal is not attained. A spherical mirror, for example, forms an image of a distant object which suffers from spherical aberration, though it is symmetrical over the field. In the Schmidt telescope the spherical aberration is corrected for the centre of the field by introducing a correcting plate having at least one non-spherical surface. This plate lies approximately in a plane passing through the centre of curvature of the mirror. The plate is only required to produce small deviations of the rays which pass through it, and these will only change slowly as the direction of the incident light is altered from a normal to oblique directions. The aberrations are thus confined to small values for a field of substantial size. The principle is applicable to a telescope comprising a number of mirrors if these are concentric spheres.

Another form of telescope which aims at the same result employs one or more weak correcting lenses in place of the Schmidt plate. These lenses are deep menisci, with only a small difference in their radii. The surfaces are spherical. This design is due to Maksutov. If both mirrors and lenses had strictly concentric spherical surfaces the problem would be solved by eliminating aberrations for a single point object. Simple examples of these two telescopes are illustrated in Figure 2.

The existence of two forms of telescope giving fields of this character inevitably raises the general question whether better correction can be secured by the use of aspherical surfaces than is possible when only plane or spherical surfaces are employed. On broad grounds we should expect the use of aspherical surfaces to be advantageous. An increase in the number of variables at our

disposal should enable aberrations to be more completely controlled in many kinds of instruments. Doubt however on this point has been cast by Herzberger, who, if I understand him aright, considers that aberrations can be corrected as fully with spherical surfaces as with any other forms. It is not clear what conditions Herzberger has assumed in reaching this conclusion. It certainly does not hold in some instruments which have a small fixed number of surfaces. But, now that it is the common practice to 'bloom' lens surfaces, modern instruments may be constructed with many more glass–air surfaces than could have been tolerated several years ago; and Herzberger may have meant that even with spherical surfaces the designer has a sufficient number of variables at his command to enable him to attain as high a state of correction as is needed.

We cannot, by appeal to Young's equations, obtain a direct answer to the question whether or not aspherical surfaces should be employed, for he has assumed that all his surfaces are spherical. But we can examine the extent to which correction is possible when only spherical surfaces are used. The simplest question for us to ask is whether more and more surfaces must be employed when

Figure 2.

the number of aberrations to be corrected is continuously increased. And by far the simplest system for us to consider is an objective giving a spherical field, constructed with a number of concentric spherical surfaces, which may be either reflecting or refracting.

The discussion of this question is facilitated by a modification of the matrix products $\mathscr{S}$ and $\mathscr{T}$. These products will not be altered if we insert between any two factors two additional factors which are the inverse of one another. For example, if the letters H and K denote non-singular matrices of the second order, the substitution

$$S'_k = K_{k-1}^{-1} S_k H_k, \qquad D'_k = H_k^{-1} D_k K_k$$

yields

$$\begin{pmatrix} \xi' & -x' \\ \eta' & -y' \end{pmatrix} K_n = \begin{pmatrix} \xi & -x \\ \eta & -y \end{pmatrix} H_0 \mathscr{S}',$$

where $\mathscr{S}' = D'_0 S'_1 D'_1 \dots S'_n D'_n$.

We choose values of H and K to obtain equations where the reference points are related to the centres of curvature of the surfaces rather than to the points of refraction. The resulting equations are for the secondary section

$$\begin{bmatrix} \dfrac{\xi'}{\zeta'} & -x'\zeta' \\[2mm] \dfrac{\eta'}{\zeta'} & -y'\zeta' \end{bmatrix} = \begin{bmatrix} \dfrac{\xi}{\zeta} & -x\zeta \\[2mm] \dfrac{\eta}{\zeta} & -y\zeta \end{bmatrix} \mathscr{S}',$$

and for the primary section

$$(\theta' - \mu'p') = (\theta \quad -\mu p)\mathscr{T}',$$

where the factors for finding $\mathscr{S}'$ and $\mathscr{T}'$ are

$$D'_k = \begin{pmatrix} \cdot & d'_k \\ 1 & 1 \end{pmatrix}, \qquad S'_k = \begin{pmatrix} 1 & \cdot \\ \sigma_k & 1 \end{pmatrix}, \qquad T'_k = \begin{pmatrix} 1 & \cdot \\ \tau_k & 1 \end{pmatrix}.$$

In these equations x, y and x', y' are the coordinates of the endpoints, which we will take to be the intersections of the incident and emergent rays with planes normal to the axis of symmetry through C_1 and C_n respectively, the centres of the first and last surfaces; and p, p' are the distances of a near ray from L_1 and L_n, the feet of the perpendiculars from C_1 and C_n to the incident and emergent rays. d'_k is $-\mu_k$ times the distance along the ray from L'_k to L_{k+1}, and

$$\sigma_k = \frac{a_k}{\zeta_{k-1}\zeta_k}, \qquad \tau_k = \frac{a_k}{\mu_{k-1}\mu_k \cos\phi_k \cos\phi'_k}.$$

We now take the special case when all the surfaces are concentric. This gives $d'_k = 0$ for every value of k, and so

$$\mathscr{S}' = \begin{pmatrix} 1 & \cdot \\ \sigma & 1 \end{pmatrix}, \qquad \mathscr{T}' = \begin{pmatrix} 1 & \cdot \\ \tau & 1 \end{pmatrix}$$

where $\sigma = \Sigma\sigma_k$ and $\tau = \Sigma\tau_k$. The order in which the surfaces are met by the light is immaterial.

The second columns of both $\mathscr{S}'$ and $\mathscr{T}'$ signify the well-known result that the product of the refractive index and the perpendicular distance of the ray from the centre of curvature is invariant. The second rows indicate that the distances of the primary and secondary principal foci from L'_n are respectively $1/\mu_n\tau$ and $1/\mu_n\sigma$. If these foci are at a constant distance from C the instrument will have a spherical image free from aberration for objects at infinity.

Consider first the special case when all the surfaces are reflectors in air. If h is the length of the central perpendicular to the ray, $h = r_k \sin\phi_k$; and by definition $\tau_k = 2/r_k \cos\phi_k$. Let us write θ_k for the power τ_k when the incidence is normal. Then $\tau_k = \theta_k(1 - \frac{1}{4}\theta_k^2 h^2)^{-\frac{1}{2}}$, and the τ condition for the position of the focus is $\tau = \theta(1 - \theta^2 h^2)^{-\frac{1}{2}}$ where $\theta = \Sigma\theta_k$. The absence of aberrations is then expressed by saying that the equation

$$\theta(1 - \theta^2 h^2)^{-\frac{1}{2}} = \Sigma\theta_k(1 - \frac{1}{4}\theta_k^2 h^2)^{-\frac{1}{2}}$$

is to be satisfied for all values of h. There are separate conditions for aberrations of different orders, and these are obtained by expanding both sides in series of ascending powers of h^2. The conditions are:

first order aberrations $\qquad \Sigma\theta_k^3 = 4\theta^3$,

second order aberrations $\quad \Sigma\theta_k^5 = 16\theta^5$,

third order aberrations $\qquad \Sigma\theta_k^7 = 64\theta^7$,

and so on.

The powers of 4 in these equations shows that aberrations of all orders are present when the objective consists of a single mirror. With two mirrors the first order aberrations can be eliminated by making

$$\frac{\theta_1}{1 + \sqrt{5}} = \frac{\theta_2}{1 - \sqrt{5}} = \frac{\theta}{2},$$

but these values give

$$\Sigma\theta_k^5 = 11\theta^5, \qquad \Sigma\theta_k^7 = 29\theta^7, \ldots.$$

so that aberrations of all higher orders are present. We can show that to correct the aberrations of the first n orders we must have at least $(n+1)$ mirrors. The conditions may be put in a very simple form. Let

$$f_n = x^n - \theta x^{n-1} + \dots.$$

be the equation satisfied by $\theta_k (k=1, 2, \dots. n)$ when the first $n-1$ aberrations are removed. The functions f for different values of n then satisfy the relation

$$f_n = x f_{n-1} - \theta^2 f_{n-2},$$

from which it is obvious that there are no zero roots of f, and that no two consecutive f's have a common root, for it would then be common to all f's. The initial values are $f_0 = 1, f_1 = x - \theta$. The number of surfaces therefore increases regularly with the number of aberrations corrected. We may state this conclusion in the form that the correction of many aberrations requires the number of degrees of freedom in the construction of the instrument to be correspondingly great.

The solutions given by these equations are not in fact of practical value. It may be shown that all the roots are real, that if n is odd there is one more positive than negative root and that otherwise positive and negative roots are equal in number. All the roots lie between the limits $\pm 2\theta$, and if arranged in order of absolute magnitude are alternately positive and negative. Under these circumstances we find that the light is inevitably on the wrong side of the mirror for it to meet the series of reflectors in the way assumed. Our temerity in reducing the design of a highly corrected optical instrument to the solution of a single algebraic equation has met with the appropriate reward. But though the solutions are worthless for the specification of real instruments, this failure does not detract from the validity of the conclusion that the removal of more aberrations can only be achieved by the utilization of a greater number of variables.

We have so far assumed that the spherically symmetrical system is composed solely of mirrors, and it is this assumption which led to the appearance of the numerical factor 4. To investigate the potentialities of objectives such as Maksutov's, which incorporate refracting as well as reflecting elements, the conditions may be expressed in a different form. Each contribution τ_k to τ consists of two terms, for

$$\tau_k = \frac{1}{r_k}\left(\frac{1}{\mu_{k-1}\cos\phi_k} - \frac{1}{\mu_k \cos\phi_k'}\right)$$

$$= \frac{\omega_{k-1}R_k}{(1-\omega_{k-1}^2 R_k^2 h^2)^{\frac{1}{2}}} - \frac{\omega_k R_k}{(1-\omega_k^2 R_k^2 h^2)^{\frac{1}{2}}},$$

where R denotes a curvature and ω the reciprocal of a refractive index. The contributions made by these terms to τ are to be taken independently. If x denotes any value of $\pm \omega R$, the conditions to be satisfied for freedom from aberrations are

$$\Sigma x_k^{2r+1} = \theta^{2r+1}, \quad (r=0,1,2\dots).$$

On the left there are two equal values of x for each reflecting surface, and for each refracting surface two values of x of opposite sign which are in a definite ratio to one another. We first form the equation of which the x's are the roots. The conditions to be satisfied by the x's are clearly satisfied by an equation which has one root equal to θ, and the other roots equal and opposite in pairs. As the total

number of roots is to be even we add one zero root. With the exception of a constant term this equation must be identical with that satisfied by our x's. The equation we require is therefore of the form

$$x(x-\theta)F+c=0,$$

where F is a polynomial in x^2. If F is of order n in x^2 the conditions $\Sigma x_k^{2r+1}=\theta^{2r+1}$ will be satisfied for all values of r from 0 to n whatever the value of c may be.

The simplest example will consist of one mirror and one lens, with air as the surrounding medium. The conditions are then of the form

$$(x-R)^2(x^2-Sx+P)(x^2+\omega Sx+\omega^2 P)$$
$$\equiv x(x-\theta)(x^4-\beta x^2+\gamma)+c,$$

and the first six terms give three equations for finding R, S, P as functions of θ and ω. This objective, consisting of three spherical surfaces, will then be free from aberrations of the first two orders.

The equation $x(x-\theta)F+c=0$ shows immediately that more than n aberrations cannot be removed if F is of order n in x^2. For if more than n aberrations could be removed the equation of the next order constructed on this pattern would have its additional roots zero, i.e. the constant c would vanish. But with θ different from zero this equation would then have an odd number of zero roots, and this is inconsistent with the relations which must subsist between pairs of roots.

The incompatibility of the restriction to a small number of surfaces and the correction of many orders of aberrations in fact became evident at an earlier stage. We found that the sums σ and τ, which relate to the positions of the sagittal and tangential foci, had to be equal for all rays. Their values are

$$\sigma=\Sigma\,\frac{1}{s_k}\left(\frac{1}{\mu_{k-1}\cos\psi_{k-1}}-\frac{1}{\mu_k\cos\psi_k}\right)$$

and

$$\tau=\Sigma\,\frac{1}{r_k}\left(\frac{1}{\mu_{k-1}\cos\phi_k}-\frac{1}{\mu_k\cos\phi'_k}\right),$$

where s_k is the projection of r_k on the z axis, and ψ_k the angle made by the ray in medium k with this axis. The condition in general is only satisfied when the surfaces are all plane—a case without interest.

From these examples then we see that, so long as we are restricted to a system of concentric spherical surfaces, the number of aberrations that can be removed depends strictly on the number of reflecting or refracting surfaces in the instrument. To obtain uniform definition over the whole of a spherical field this condition of concentricity is essential. Since perfect correction is out of the question a compromise is unavoidable, and its nature will depend on the purposes for which the instrument is to be used. We may for instance depart from the condition that only spherical surfaces are to be used, as Schmidt has done, or abandon the condition of concentricity after the example of Maksutov. By so doing we increase the number of variables we can employ to control aberrations, and in favourable conditions may thereby enlarge the aperture or field of view or both while maintaining a high standard of definition. An immense amount of labour has been spent in attempts to improve the performance of instruments in which the surfaces are all spherical, and the results so far attained fall short of our desires. We do not know whether further substantial improvements are possible under

these conditions: but the conditions which enable improvements to be attained are certainly not readily found. On the evidence so far known it appears that aspherical surfaces may well prove of great importance in the further development of optical instruments, and I should like to be more certain than I am now that this is appreciated by our manufacturers as well as by our optical designers and engineers.

I have approached these questions from the point of view of Thomas Young, and have necessarily discussed them on the assumption that the surfaces of the instrument are spherical. More powerful methods not subject to this limitation are known, though at present they are not very extensively used. If the views to which I have been led are correct these more powerful methods are likely to replace those which depend on the sphericity of instrumental surfaces. In a lecture which should rather eulogize Young's scientific contributions I have been led to suggest that they will be of less assistance to our successors than they have been to ourselves. Perhaps I ought to apologize for such a misdemeanour. But the real testimony to the importance of Young's work in the field of geometrical optics is the fact that his equations are still used by optical designers one hundred and fifty years after he obtained them.

The Diffuse Reflectivity of Oxide Layers Formed on Steels Under Defined Conditions

By R. J. DONATO

South-West Essex Technical College, London

ABSTRACT. A method to determine the reflectivity of oxide layers formed on steels under defined conditions at various wavelengths in the visible spectrum and at different temperatures up to 660°c. is described. An integrating sphere is used in which the specimen replaces a portion of the sphere wall, and a method is given for the simultaneous comparison of two beams of light using a single photocathode. Except for a correction factor depending upon the geometry of the sphere, the measurements rely solely upon the area of an adjustable aperture. Some results are given for the oxides of various steels.

§1. INTRODUCTION

THE determination of the reflectivity of iron and steel oxides is important from both theoretical and practical considerations. The connection between the reflectivity and the optical constants enables the position of absorption bands to be detected, and any data upon the reflectivity at different temperatures and wavelengths are of service to the theory of solids. From the practical standpoint the values of reflectivity obtained for iron and steel oxides are almost directly applicable to optical pyrometry through the Kirchhoff relationship

$$E_\lambda + r_\lambda = 1, \qquad \ldots\ldots(1)$$

where E_λ and r_λ represent the emissivity and reflectivity at wavelength λ, both being measured at the same temperature. Clearly, to be of any use at all in this connection, the reflectivity readings must be taken at different temperatures, and, if information applicable to two colour pyrometers is to be obtained, at various wavelengths.

As the oxides of iron and steel present rough surfaces a specular reflection method is impossible. Accordingly in the present method an integrating sphere was used, and the diffuse reflectivity obtained as a ratio of the brightness of the oxide to the brightness of the wall of the integrator. Diffuse reflectivity methods have been used by Hamaker (1934), Ornstein and van der Deen (1936), and Prescott and Morrison (1939), but in all these methods two separate measurements were made, one upon the substance of unknown reflectivity and the other upon a standard surface. By a simultaneous comparison, however, any variation in the light sources would affect both beams equally and so remove any errors that might occur by taking two separate measurements.

§2. THEORY

One of the first exponents of the theory of the integrating sphere was Sumpner (1892). He stated that if a sphere is coated internally with a perfect diffuser possessing little or no absorption and a light source placed within, then the brightness of any part of the sphere due to light reflected from the remainder will be the same and proportional to the total flux emitted from the source. Further, if an amount of light flux F lumens entering the sphere serves to illuminate a small area of the wall, of reflectivity ρ, then an amount of light $F\rho$ will be reflected from the illuminated area and after a large number of further reflections will be absorbed.

Thus
$$F\rho = IA(1-\rho), \qquad \ldots\ldots(2)$$

where I denotes the illumination within the sphere, A the surface area, and $(1-\rho)$ the absorption coefficient.

Then
$$I = F\rho/A(1-\rho). \qquad \ldots\ldots(3)$$

If the specimen of unknown reflectivity is made to occupy a portion of the sphere wall and a screen is inserted to prevent direct light from the illuminated area reaching the specimen (see X and Y in Figure 1), the illumination at the specimen will be reduced to

$$i = F\rho/A(1-\rho) - F\rho/A = \frac{F\rho}{A}\frac{\rho}{1-\rho}. \qquad \ldots\ldots(4)$$

Comparison between the brightness of the specimen ($b=ri$, where r is the reflectivity of the specimen) and the brightness of a part of the sphere wall ($B=\rho I$) yields

$$\frac{b}{B} = \frac{ri}{\rho I} = \frac{r\dfrac{F\rho}{A}\dfrac{\rho}{1-\rho}}{\rho F\rho/A(1-\rho)} = r. \qquad \ldots\ldots(5)$$

Thus the ratio of the brightness of the specimen to the brightness of a portion of the whitened sphere wall gives an absolute measure of the specimen reflectivity. If the specimen is viewed normally as in Figure 1, then since its surface is not a perfect diffuser the reflectivity obtained by equation (5) is not the total reflectivity, in the sense that it is the ratio of the total amount of reflected light to the total amount of completely diffused incident light, but is the total reflectivity for light incident unidirectionally and normal to the surface. This has been shown to be so by McNicholas (1928) who obtained by the Helmholtz reciprocity theorem a connection between the apparent reflectivity for completely diffused incident illumination with observation normal to the surface (defined as the ratio $\pi B'/\pi B_0$,

where B' is the light reflected normal to the surface and πB_0 the amount of incident flux, in this case completely diffused) and the total reflectivity for light incident normal to the surface. Hence the reflectivity measured is directly connected with the normal emissivity (i.e. the emissivity measured normal to the surface) by the Kirchhoff relationship. As this is the emissivity usually measured in
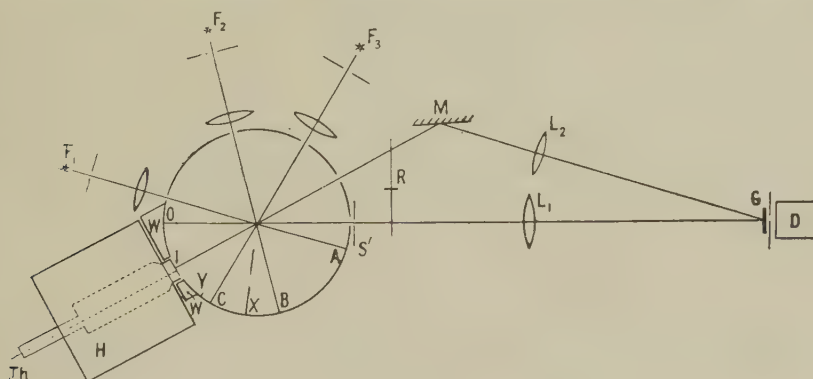
Figure 1.

M=Mirror.	H=Furnace.
R=Rotating sector.	I=Specimen.
S′=Adjustable aperture.	W=Water-cooling jacket.
L₁, L₂=Biconvex lenses, 25 cm. focal length.	Th=Thermocouple.
F₁, F₂, F₃=500 watt 110 volt lamps.	O=Observed spot of sphere wall.
A, B, C=Illuminated areas of sphere wall.	G=Filter.
X, Y=Screens.	D=Detector.

practice, the reflectivity readings made with the present apparatus may be directly converted into normal emissivities and used in optical pyrometry.

The principle of the present method then lies in comparing the intensities of two beams of light, one from the specimen and the other from a portion of the sphere wall.

§3. METHOD OF COMPARISON

When two square wave-forms in phase opposition are superimposed, the resultant is itself a square wave except when the two square waves are of equal amplitude; it is then a straight line. This is illustrated in Figure 2. Thus if a beam of light is so interrupted that a graph of light intensity against time has a periodic square wave-form, and to this beam another is added having a similar wave-form but π out of phase with the first, the resultant will be either uniform intensity or another square wave according to whether the two beams are equal or not in intensity. This addition of two wave-forms in phase opposition formed the basis of the method of detection, and adjustments were made until the square waves were equal in magnitude (Figure 2(c)). Such a method had the advantage that the two beams emerging from the sphere are viewed simultaneously, thereby making the comparison independent of fluctuations in the lamps used to supply the incident illumination. Furthermore, if a single photocathode is used to transmute variations of light intensity into current variations, the same part of the cathode may be used by both beams and so obviate errors that might arise through using different parts of the photocathode.

To give the two beams emerging from the sphere a periodic square-wave pattern, a rotating sector was used. The sector was 18 cm. in outer diameter

and had two right-angle spacings in it diametrically opposite to each other. The positions of the two beams in relation to the sector are shown in Figure 3 by B_1 and B_2. As long as the cross-sectional areas of the beams at the sector were small compared with the sector opening, the two beams would follow a square-wave variation; should the area of the beam be not small compared with the sector opening, then the resultant would have a wave-form with a flat top but with the vertical parts of the wave rounded towards the top. In actual fact it was found difficult to get the exact out of phase relationship between the two beams, and this effect coupled with the rounding of the vertical parts of the wave served to produce the wave trace shown in Figure 2(d). Nevertheless, these peaks, as long as they were not too broad, did not disturb the viewing of the balance point when the wave trace was finally obtained on a cathode-ray oscillograph; in fact, they helped

(a) Wave form of single beam

(b) Resultant of two unequal beams

(c) Resultant of two equal beams

(d) Wave trace in practice when two beams are equal
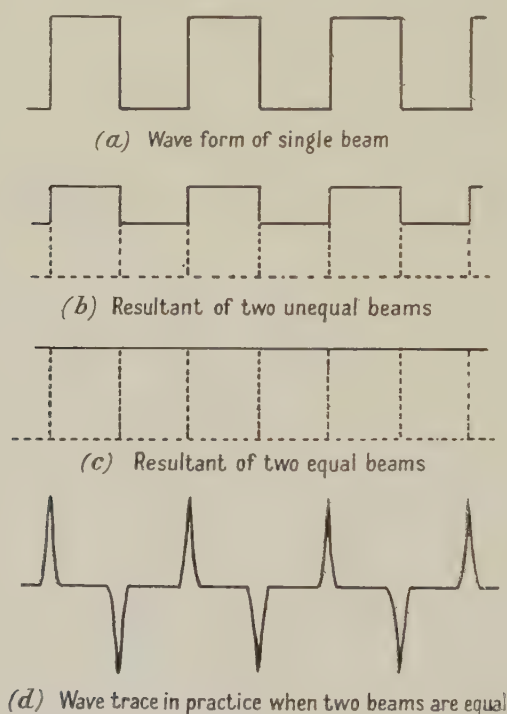
Figure 2.

to synchronize the trace. The two beams were judged to be equal in magnitude when the horizontal parts of the trace were in a straight line, as shown in Figure 2(d).

In order that the two beams emerging from the sphere should produce square waves of equal magnitude, an adjustable aperture was placed in the beam having the larger intensity, viz. the beam from observed spot O of the sphere wall (termed the standard beam). By suitable adjustment of the aperture the quantity of light in the standard and specimen beams could be made equal. Measurements were made with (i) the specimen replaced by a portion of the sphere wall (magnesium oxide) and the screens X and Y removed from the sphere, (ii) the magnesium oxide still in position but with the screens replaced, and (iii) the specimen (reflectivity r) re-introduced and the screens retained.

Let k be a constant depending upon the geometry of the optical system, e.g. different optical paths, transmissions, solid angles etc. in the two beams, S the area of the aperture in the specimen beam (kept constant as it is a hole in the sphere), and let I, with the appropriate subscripts, denote the intensities in each beam, and S', with the appropriate subscripts, the area of the adjustable aperture; then

$$\text{(i)} \qquad\qquad kI_0S = I_0S'_0, \qquad\qquad\qquad \dots\dots(6)$$

$$\text{(ii)} \qquad\qquad kI_\varrho S = I'_\varrho S'_\varrho, \qquad\qquad\qquad \dots\dots(7)$$

$$\text{(iii)} \qquad\qquad kI_r S = = I'_r S'_r. \qquad\qquad\qquad \dots\dots(8)$$

Hence

$$\text{(ii)} \qquad\qquad \frac{I_\varrho}{I'_\varrho} = \left(\frac{b}{B}\right)_\varrho = \frac{S'_\varrho}{kS} \qquad\qquad \dots\dots(9)$$

and

$$\text{(iii)} \qquad\qquad \frac{I_r}{I'_r} = \left(\frac{b}{B}\right)_r = \frac{S'_r}{kS}. \qquad\qquad \dots\dots(10)$$

Hardy and Pineo (1931) showed that if the area of the specimen is a, the area of the holes in the sphere α, and the inside area of the sphere surface $A'(=A-a-\alpha)$,
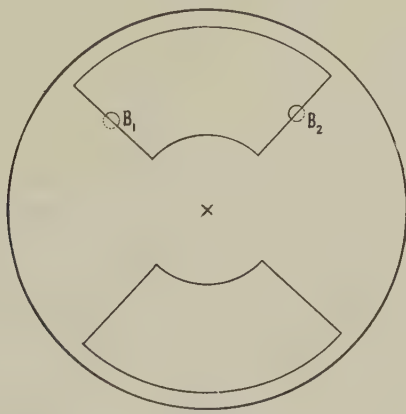
Figure 3.

then equation (5) would have to contain a correction factor. The expressions in the two cases then become

$$\left(\frac{b}{B}\right)_r = r\left\{1 - \frac{\alpha + (1-r)a}{\rho A'}\right\} = \frac{S'_r}{kS}, \qquad\qquad \dots\dots(11)$$

$$\left(\frac{b}{B}\right)_\varrho = \rho\left\{1 - \frac{\alpha + (1-\rho)a}{\rho A'}\right\} = \frac{S'_\varrho}{kS}. \qquad\qquad \dots\dots(12)$$

The quantities actually measured in the present method were

$$\left(\frac{b}{B}\right)_r \Big/ \left(\frac{b}{B}\right)_\varrho \qquad\text{and}\qquad \left(\frac{b}{B}\right)_\varrho,$$

and combining the expressions with equation (6)

$$\left(\frac{b}{B}\right)_r \Big/ \left(\frac{b}{B}\right)_\varrho = \frac{r}{\rho}\left\{1 - \frac{(\rho - r)a}{\rho A'}\right\} = \frac{S'_r}{S'_\varrho}, \qquad\qquad \dots\dots(13)$$

$$\left(\frac{b}{B}\right)_\varrho = \rho\left\{1 - \frac{\alpha + (1-\rho)a}{\rho A'}\right\} = \frac{S'_\varrho}{S'_0}, \qquad\qquad \dots\dots(14)$$

from which two expressions r may be calculated.   Thus apart from being inde-
pendent of lamp fluctuations and variations in the photocathode, the method
does not depend upon the geometry of the optical system as k does not occur in
equations (13) and (14).   A further correction has to be applied for the different
spacings, relative to the specimen and spot O, of the illuminated areas A, B and
C (Figure 1), and its value has been worked out by McNicholas (1928).   This
correction factor is by no means certain as it assumes A, B and C to be equal in
all respects.   Finally, Taylor (1935) found that the error caused by the screens
influencing the spherical conditions could be minimized if the screens were
whitened and their surface area made small compared with the area of the sphere
wall.   In such a way he found the error could be reduced to less than 0·5%.

§4. APPARATUS AND EXPERIMENTAL PROCEDURE
(i) *The Sphere*

The sphere, having an internal diameter of $9\frac{7}{8}$ in., consisted of two hollow
copper hemispheres spun from sheet of thickness 18 s.w.g. fitting together by a
bayonet pin and socket arrangement.   Three holes of diameter 1·9 cm. were
drilled on a great circle of the sphere and allowed light from three 500 watt,
110 volt (D.C.) lamps, F_1 and F_2 and F_3, to illuminate the sphere wall at A, B and C
(Figure 1).   As a rotating sector was being used to modulate the light beams
with a square wave-form, the incident illumination had to originate from a D.C.
source.   This method of using three incident beams is due to McNicholas (1928)
who found that the distribution of illumination over the sphere compared favour-
ably with perfect spherical conditions.   The angles between F_1 and F_2, F_2 and
F_3 were 60° and 45° respectively.   Two more holes, this time of diameter 1·35 cm.,
were drilled on the same great circle, one diametrically opposite the specimen
aperture (see below) and the other subtending an angle of 28° at the centre of the
sphere with the first.   This second hole was diametrically opposite the observed
spot O.   The size of all these holes was determined by two conflicting consider-
ations—the attainment of the greatest amount of light without destroying too
much the spherical distribution.   Finally the specimen hole, having a diameter
of $1\frac{1}{8}$ in., was also drilled on the great circle and made large enough to contain
the specimens.   The specimen hole was so fixed that the specimen beam was
inclined at 44° to the incident beam from F_1.

Two rectangular screens, X and Y, of dimensions $3\frac{1}{4}$ in. × 1 in., $\frac{3}{16}$ in. × 1 in.
were fitted through two slits 1 in. long perpendicular to the plane of the circle
containing the holes.   The largest screen was that farthest from the specimen
aperture.   As the function of the screens was to prevent direct light from A, B
and C reaching the specimen but at the same time not to intercept the direct light
to O, screen X had a rectangular portion cut from it to allow the direct light to
pass from B to O.

Around the specimen aperture a brass water cooling jacket was soldered to
the outside of the sphere.   Due to the presence of screens X and Y the jacket
was fitted asymmetrically with the greater extension in the direction of observed
spot O, the direction in which the cooling was most important.

The inside surface of the sphere and the screens were coated with white
diffusing magnesium oxide formed by burning magnesium ribbon in a Kodak
burner.   To form a foundation to the coating, the sphere and screens were dull
chromium plated and white paint was applied to those portions where chipping

of the magnesium oxide might have occurred, viz. around the holes and the joint between the hemispheres. In this way a reasonably diffusing surface would have been presented even if flaking had occurred. The hemispheres were coated separately.

Brass collars, which served as lens holders, were waxed round the incident and emergent holes. The sphere was clamped firmly between two horizontal squares of wood (11 in. × 11 in. × ½ in.), each having a hole of diameter 4 in. cut from their centres. Four rods of iron screwed into the corners of the squares supplied vertical supports and the sphere was adjusted so that all the holes were at the same height. The adjustable aperture, a hole cut in a stout piece of brass with adjustments made by micrometer movements of a straight edge across the hole, was screwed to the wooden framework so that it fell immediately adjacent to the hole opposite spot O. The advantage of this shaped aperture was that it was more sensitive at smaller openings, i.e. at smaller intensities.

(ii) *The Furnace*

A solid cylinder of iron, 5½ in. long, 1¾ in. diameter, was drilled down the centre through the length of the iron. This cylindrical cavity of ¼ in. diameter served to carry a thermocouple sheathed in silica tubing. The specimen end of the cylindrical cavity was tapped and a hollow screw fitted protruding from the end face. The specimens, discs 1 in. diameter, ⅜ in. thick, were drilled centrally from the back surface and tapped. Drilling was performed until the hole nearly reached the front oxide surface, that is, so that the thermocouple could be placed to within a few thousandths of an inch of the surface. As the specimens were tapped it was possible to screw them on to the hollow screw until their back surface made contact with the machined face of the cylinder. More uniform heating across the specimen was assured by having the cylinder of larger diameter than the specimen. This end of the cylinder was bevelled to make it fit easily into the containing box which enabled the cylinder to be packed tightly with Kieselguhr.

The furnace was heated electrically by nichrome wire wound round the cylinder and insulated from it by a layer of Pyruma cement. The other end of the furnace was supplied with a grip for easier handling formed by screwing a piece of conduit tubing to the furnace. A three-legged stand fitted with levelling screws supported the furnace container, and the specimen was manœuvred into the aperture by hand. No appreciable errors were introduced this way as the reflectivity measured normal to the surface varies only very slightly with the angle in the range normal to 15° from normal. Temperatures were measured by means of a 13% platinum–platinum-rhodium thermocouple and a Tinsley potentiometer. The furnace is outlined and marked H in Figure 1.

(iii) *The Optical System*

The beams from sources F_1, F_2 and F_3 were focused by means of biconvex lenses, focal length 6 cm., to give images of the filaments at A, B and C, and an image of spot O was formed at the detector aperture by means of a biconvex lens L_1. As the adjustable aperture S′ was in the beam from spot O (an almost perfect diffuser) the intensity of light was constant over the area S′ to a high degree of accuracy. However, in order to make the area of S′ directly proportional to the light passing through it, S′ had to be the limiting aperture of the optical

system, and with this condition in mind the optical system was designed. The specimen beam was deviated by means of a front-aluminized mirror M and passed through a biconvex lens L_2. Two superimposed images were thus obtained at the detector aperture, and by suitably placing the rotating sector the phase opposition obtained. The detector aperture was designed so that its area was smaller than the image of the specimen. In this way light was not collected from the extreme edges of the specimen where the temperature was found to fall off.

To select different wavelengths in the visible spectrum four Ilford monochromatic filters were placed in turn immediately in front of the detector aperture. These corresponded to maximum transmission at wavelengths 0.43μ, 0.47μ, 0.49μ, 0.52μ and with transmission over wavelength regions of 0.09μ, 0.05μ, 0.045μ and 0.045μ respectively.

(iv) The Comparator

A photomultiplier (RCA 931 A) having an antimony–cæsium cathode with a cut-off at 0.7μ was used to receive the two beams. The over-all amplification of the cell was about 10^5 and the output was connected to a cathode-ray oscillograph. Observations were most satisfactory when the rotating sector was running at approximately 25 cycles per second. The noise produced and subsequently amplified by the cathode-ray tube amplifier was found not to be in excess, and the noise-to-signal ratio provided a determining factor in the size of the holes on the sphere.

(v) Experimental Procedure

With the particular wavelength and temperature range (room temperature to $660°$ c.) no correction was found to be necessary for self-luminosity.

The specimens were heated for about three hours in air to a temperature of $660°$ c. and readings taken on cooling. So long as the oxide is formed at temperatures above $600°$ c. analysis shows the composition of the oxide layer on carbon steels is independent of time (Pfeil 1929) and consists of a mixture of FeO and Fe_2O_3. Readings were taken at various temperatures by adjusting the aperture in the standard beam until a wave trace similar to that in Figure 2 (d) was obtained.

A correction was applied to take account of the fact that the front surface temperature of the oxide was less than the temperature recorded by the thermocouple. The method of microscopic melts was used (Burgess and Foote 1915) and the substances used were tin and sulphur. Finally a reading was taken with an optical pyrometer at a temperature above $700°$ c., a value being assumed for the oxide emissivity. All these measurements indicated that the surface temperature was about $20°$ less than the thermocouple reading.

Readings were taken with the specimen replaced by a sample of magnesium oxide with and without the screens X and Y (equations (13) and (14)).

(vi) Results

The curves of reflectivity against temperature for three different types of iron and steel are shown in Figures 4, 5 and 6. They are for 0.04% C, 0.45% C and 13% Cr. A more detailed analysis is given below.

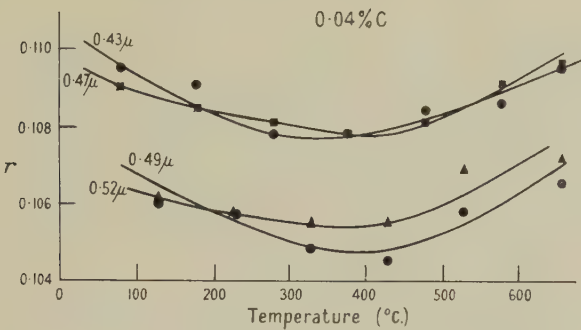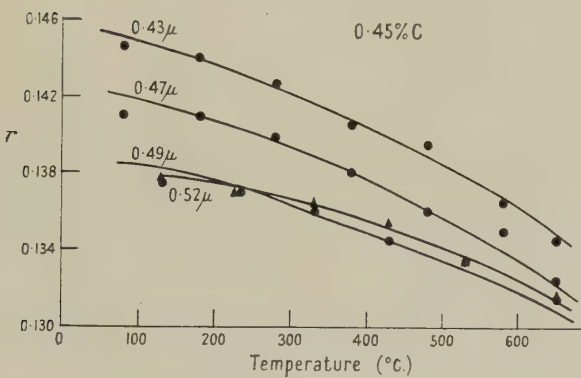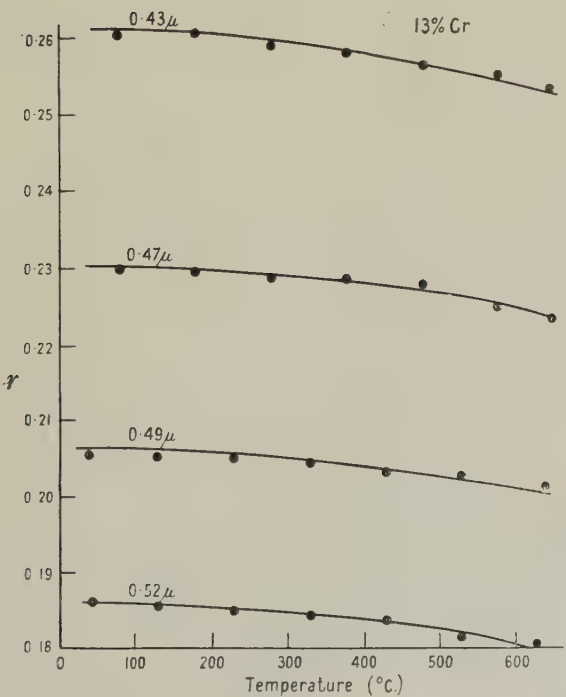|	C	Mn	Si	S	P	Ni	Cr	Cu
0·04% C	0·04	0·04	0·05	0·011	0·027	0·04	0·01	0·05
0·45% C	0·45	0·66	0·18	—	—	—	—	—
13% Cr	0·07	0·35	0·31	0·017	0·017	—	12·66	—

Figure 4.

Figure 5.

Figure 6.

All the samples were provided by The United Steel Companies Ltd., Sheffield, and were previously oxidized according to the specification: $0.04 C\%$—1 hour at $600°$ c.; $0.45\% C$—1 hour at $600°$ c.; $13\% Cr$—3 hours at $750°$ c. For the first two types of specimen the oxide was greyish-black and showed no tendency to flake on cooling—probably due to the cold working treatment the specimens received during preparation. The $13\% Cr$ presented a shiny surface in spite of its oxidation and it exhibited a blue film. The state of this surface remained unchanged on further heating; this probably explained the fact that the reflectivity was greater at the shorter wavelengths.

Several heatings were carried out with each specimen and a set of curves obtained parallel to the typical ones shown in Figures 4, 5 and 6. Considering this spread the calculated error is 1.5% or less.

The curves show that the actual value of reflectivity varies not only with the composition of the iron, but also with temperature. The negative slopes in Figures 5 and 6 might be explained on the theory of Weil (1948) if it can be assumed that the semiconducting properties of the oxides play a predominant rôle, for the sign of the variation of conductivity of semiconductors with temperature is opposite to that of pure metals. Thus on this free electron theory it would seem that the temperature variation of reflectivity in the visible would also be opposite to that of pure metals. It is far more likely, however, that the effect of the dielectric properties of the oxide have to be considered. Comparison with results of previous workers is rendered difficult as nearly all of them measured the emissivity or reflectivity over the whole wavelength range (e.g. Randolph and Overholser 1913). The general view is that this variation of reflectivity with temperature is negative, and inasmuch as the greater part of the energy is in the infra-red it appears that results based upon a consideration of conductivity alone are incorrect.

The minima obtained for the $0.04\% C$ may indicate the presence of an absorption band whose position depends upon the temperature. Any further analysis at this stage is impossible because of the number of factors governing the behaviour of bound electrons.

REFERENCES

BURGESS, G. K., and FOOTE, P. D., 1915, *Bur. Stand. Bull., Wash.,* **12**, 83.
HAMAKER, H. C., 1934, *Doctor's Thesis,* Utrecht.
HARDY, A. C., and PINEO, O. W., 1931, *J. Opt. Soc. Amer.,* **21**, 502.
McNICHOLAS, H. J., 1928, *Bur. Stand. J. Res., Wash.,* **1**, 29.
ORNSTEIN, L. S., and van der DEEN, J. H., 1936, *Physica,* **3**, 289.
PFEIL, L. B., 1929, *J. Iron and Steel Inst.,* **119**, 501.
PRESCOTT, C. H., and MORRISON, J., 1939, *Rev. Sci. Instrum.,* **10**, 36.
RANDOLPH, C. P., and OVERHOLSER, M. J., 1913, *Phys. Rev.,* **2**, 144.
SUMPNER, W. E., 1892, *Proc. Phys. Soc.,* **12**, 10.
TAYLOR, A. H., 1935, *J. Opt. Soc. Amer.,* **25**, 51.
WEIL, R. A., 1948, *Proc. Phys. Soc.,* **60**, 8.

Transit Time Correction Factor for Cylindrical Noise Diodes

By H. ASHCROFT * and C. HURST †

* The Queen's College, Oxford   † Jesus College, Oxford

ABSTRACT. The noise generated by a diode is always less than it would be if the transit of the electrons were instantaneous. The factor by which it is reduced has been evaluated for transit angles up to $12\frac{1}{2}$ radians and for ratios of anode to cathode radius up to 20. It has been found to fall to zero for one pair of values of these parameters.

§1. INTRODUCTION

WHEN electrons leave the cathode of a diode, the corresponding current in the outside circuit connecting anode and cathode flows, not instantly, but gradually during the transit of the electrons to the anode. For, when the electrons leave the cathode, an equal induced positive charge is formed on the cathode surface behind them, and, as they approach the anode, this charge diminishes while an induced positive charge on the anode grows. When the electrons reach the anode, the charge induced by them on the cathode has fallen to zero and the induced charge on the anode is just neutralized by the arrival of the electrons. The circuit current is merely the transfer, which has been taking place continuously during the transit, of induced charge from one electrode to the other. The effect may be summed up quantitatively in the statement that, when a charge q moves in the vicinity of a system of conductors, then the instantaneous current induced in one of the conductors is qvE_v; here, v is the velocity of the charge and E_v is the component of electric field in the direction of v which would exist if the charge itself were removed and if the conductor under consideration were raised to unit potential, the other conductors being grounded. This important theorem is usually ascribed to Ramo, who enunciated it in 1939; but, in fact, it was formulated by Shockley a year earlier (Shockley 1938).

From a knowledge of the variation of vE_v during an electron's flight across a particular diode, one could determine the shape of the current pulse to which its emission and transit give rise. If the Fourier integral of this pulse were found and if the components were summed for all electrons in the phases appropriate to the random succession of their emission, one could find the amount of circuit current—or noise—in any band of frequencies. An equivalent and more convenient derivation of the noise power at any frequency may be obtained, however, by imagining sinusoidal variations at that frequency and of the amplitude given by Schottky's equation to be superimposed on the steady emission. If this extra current is $i\cos\omega t$ and if the current pulse (i.e. vE_v) for the passage of a unit charge is $\phi(\tau)$ (τ = time from emission), then, at time t, the charge emitted in an interval $\delta\tau$ at a time τ ago was $i\cos\omega(t-\tau)\delta\tau$. Thus the whole circuit current induced is $i\int_0^{} \phi(\tau)\cos\omega(t-\tau)\,d\tau$, the upper limit of integration being the time taken for an electron to reach the anode. This current may be resolved into two components,

one in phase and the other in quadrature with the emission current. These are, respectively,

$$i \cos \omega t \int_0 \phi(\tau) \cos \omega \tau \, d\tau \quad \text{and} \quad i \sin \omega t \int_0 \phi(\tau) \sin \omega \tau \, d\tau.$$

Any calculation of noise power made by assuming that the circuit current fluctuations equal those in the emission can therefore be corrected by multiplying by the factor

$$\left[\left\{ \int_0 \phi(\tau) \cos \omega \tau \, d\tau \right\}^2 + \left\{ \int_0 \phi(\tau) \sin \omega \tau \, d\tau \right\}^2 \right].$$

We call this quantity the transit time correction factor.

In the present computations no account is taken of the effects of the mutual repulsion of the electrons or of their emission velocities. The simplifications are justified since diode noise generators making use of shot noise are run under temperature limited conditions and with anode voltages great enough to make the velocities of flight much greater than those of emission (Kompfner, Hatton, Schneider and Dresel 1946). Noise due to secondary electrons and to electrons which have been reflected from the anode is also neglected.

If, with these assumptions, V_r is the potential and v_r the velocity at radius r in a cylindrical diode and suffixes $_k$ and $_a$ are used to denote cathode and anode values, then $v_r^2 = 2(e/m)V_r$ and the field which would exist if the cathode were grounded and the anode raised to unit potential is $(1/V_a)(dV_r/dr)$. The current induced in the circuit by a moving electron is therefore

$$e \frac{1}{V_a} \frac{dV_r}{dr} v_r = \frac{e}{V_a} \left(\frac{2V_r e}{m} \right)^{\frac{1}{2}} \frac{dV_r}{dr}.$$

The time, τ, to reach radius r is given by

$$\tau = \int_{r_k}^{r} v^{-1} \, dr = (m/2e)^{\frac{1}{2}} \int_{r_k}^{r} V^{-\frac{1}{2}} \, dr.$$

Hence, by putting $u = \ln(r/r_k)$ and noting that $V_r = V_a u/u_a$, we get

$$\tau = (m/2e)^{\frac{1}{2}} r_k (u_a/V_a)^{\frac{1}{2}} \int_0^u u^{-\frac{1}{2}} e^u \, du.$$

Thus $\omega \tau = \theta I/I_a$ if I denotes $\int_0^{\sqrt{u}} \exp(x^2) dx$ and if θ denotes

$$\omega (m/2e)^{\frac{1}{2}} r_k (u_a/V_a)^{\frac{1}{2}} \int_0^{u_a} u^{-\frac{1}{2}} e^u \, du,$$

the whole transit angle. It follows that

$$\phi(\tau) = \frac{1}{V_a} \left(\frac{2V_r e}{m} \right)^{\frac{1}{2}} \frac{dV_r}{dr}$$

and that $\delta\tau = (2V_r e/m)^{-\frac{1}{2}} \delta r$, so that $\phi(\tau)\delta\tau = V_a^{-1}\delta V_r = u_a^{-1}\delta u$.

The transit time correction factor is therefore

$$\left[\left\{ u_a^{-1} \int_0^{u_a} \cos(\theta I/I_a) \, du \right\}^2 + \left\{ u_a^{-1} \int_0^{u_a} \sin(\theta I/I_a) \, du \right\}^2 \right]. \quad \ldots\ldots(1)$$

§2. EXPLANATION OF THE TABLES

The derivation of the above expression for the transit time correction factor has been given by Spenke (1937) who plotted the square root of the factor for $u_a = 0$ (when the factor reduces to $\{(\theta - \sin\theta)^2 + (1 - \cos\theta)^2\}/\frac{1}{4}\theta^2$) and for $u_a = 1.5$

and 4 between $\theta=0$ and 12 radians. The object of the present work was the tabulation of the factor at intervals of u_a sufficiently fine to cover the range $u_a=0$ to 3. The method of computation is described in the later sections of this paper. The procedure for using the tables is:

(i) Find u_a, the logarithm to base e of the ratio of anode radius to cathode radius.

(ii) Find θ, the transit angle.

$$\theta = \omega(2e/m)^{-\frac{1}{2}}r_k(u_a/V_a)^{\frac{1}{2}}\int_0^{u_a} u^{-\frac{1}{2}}\exp u\, du$$

$$= 2\omega d(2V_a e/m)^{-\frac{1}{2}}\{u_a^{\frac{1}{2}}I_a(\exp u_a-1)^{-1}\},$$

where ω is 2π times the frequency, d the distance separating anode and cathode and V_a the applied anode voltage. The function $\{u_a^{\frac{1}{2}}I_a(\exp u_a-1)^{-1}\}$ of u_a is tabulated in Table 1.

(iii) Enter Table 2. The factor given there is the ratio of circuit current fluctuations squared to emission current fluctuations squared. If the fourth decimal place is not required linear interpolation is sufficient in the rows and, beyond about 6 radians, in the columns. Otherwise, second differences have usually to be taken into account; third differences can always be neglected.

Table 1

u_a	$\dfrac{I_a\sqrt{u_a}}{\exp u_a-1}$	u_a	$\dfrac{I_a\sqrt{u_a}}{\exp u_a-1}$	u_a	$\dfrac{I_a\sqrt{u_a}}{\exp u_a-1}$	u_a	$\dfrac{I_a\sqrt{u_a}}{\exp u_a-1}$
0·0	1·000	0·8	0·878	1·6	0·780	2·4	0·706
0·2	0·967	1·0	0·851	1·8	0·759	2·6	0·691
0·4	0·936	1·2	0·826	2·0	0·740	2·8	0·677
0·6	0·906	1·4	0·802	2·2	0·722	3·0	0·664

§ 3. THE METHOD OF COMPUTATION

To produce the tables, expression (1) was first converted into a power series in θ, the coefficients being functions of u_a. One has:

$$\frac{1}{u_a}\int_0^{u_a}\cos\frac{\theta I}{I_a}\,du = \sum_{r=0}^{\infty}\frac{(-1)^r\,\theta^{2r}}{I_a^{2r}u_a(2r)!}\int_0^{u_a} I^{2r}du, \quad \text{on expanding the cosine,}$$

$$= \sum_{r=0}^{\infty}\frac{(-1)^r}{(2r)!}J_{2r}\theta^{2r}, \quad \text{where} \quad J_n \equiv \frac{1}{u_a I_a^n}\int_0^{u_a} I^n\,du.$$

J_n is a function of u_a which never exceeds unity and thus the latter series for $u_a^{-1}\int_0^{u_a}\cos(\theta I/I_a)\,du$ is (by comparison with $\cosh\theta$) absolutely convergent for any assigned value of θ. It may therefore be squared by being multiplied by itself, term by term. Therefore,

$$\left\{\frac{1}{u_a}\int_0^{u_a}\cos\frac{\theta I}{I_a}\,du\right\}^2 = \sum_{r=0}^{\infty}\left\{\frac{(-1)^r}{(2r)!}\,\theta^{2r}\sum_{s=0}^{r}\binom{2r}{2s}J_{2s}J_{2r-2s}\right\}.$$

Similarly $\dfrac{1}{u_a}\displaystyle\int_0^{u_a}\sin\frac{\theta I}{I_a}\,du = \sum_{r=0}^{\infty}\frac{(-1)^r}{(2r+1)!}J_{2r+1}\theta^{2r+1}$

and $\left\{\dfrac{1}{u_a}\displaystyle\int_0^{u_a}\sin\frac{\theta I}{I_a}\,du\right\}^2 = \sum_{r=0}^{\infty}\left\{\frac{(-1)^r}{(2r)!}\,\theta^{2r}\sum_{s=0}^{r-1}\binom{2r}{2s+1}J_{2s+1}J_{2r-2s-1}\right\}.$

Table 2.   Transit Time

θ radians	u	0·00	0·20	0·40	0·60	0·80	1·00	1·20
	R	1·000	1·221	1·492	1·822	2·226	2·718	3·320
0·0		1·0000	1·0000	·1·0000	1·0000	1·0000	1·0000	1·0000
0·25		0·9965	0·9964	0·9964	0·9963	0·9962	0·9961	0·9960
0·5		0·9862	0·9858	0·9855	0·9851	0·9848	0·9845	0·9842
0·75		0·9692	0·9684	0·9676	0·9668	0·9661	0·9654	0·9648
1·0		0·9458	0·9444	0·9430	0·9417	0·9404	0·9392	0·9381
1·25		0·9165	0·9143	0·9122	0·9102	0·9082	0·9064	0·9047
1·5		0·8818	0·8787	0·8757	0·8729	0·8701	0·8676	0·8653
1·75		0·8423	0·8383	0·8342	0·8304	0·8269	0·8235	0·8205
2·0		0·7988	0·7936	0·7885	0·7837	0·7792	0·7750	0·7712
2·25		0·7519	0·7455	0·7393	0·7335	0·7280	0·7229	0·7183
2·5		0·7025	0·6949	0·6875	0·6806	0·6741	0·6681	0·6627
2·75		0·6513	0·6425	0·6339	0·6259	0·6184	0·6115	0·6054
3·0		0·5992	0·5891	0·5794	0·5703	0·5618	0·5541	0·5473
3·25		0·5469	0·5356	0·5248	0·5146	0·5052	0·4968	0·4894
3·5		0·4952	0·4828	0·4709	0·4597	0·4495	0·4404	0·4325
3·75		0·4448	0·4312	0·4183	0·4063	0·3954	0·3858	0·3774
4·0		0·3963	0·3817	0·3679	0·3551	0·3436	0·3335	0·3248
4·25		0·3502	0·3347	0·3201	0·3067	0·2947	0·2843	0·2754
4·5		0·3070	0·2907	0·2755	0·2616	0·2492	0·2385	0·2297
4·75		0·2670	0·2501	0·2344	0·2201	0·2075	0·1967	0·1880
5·0		0·2305	0·2131	0·1970	0·1825	0·1698	0·1591	0·1506
5·25		0·1977	0·1799	0·1636	0·1490	0·1363	0·1258	0·1177
5·5		0·1687	0·1507	0·1342	0·1197	0·1072	0·0970	0·0893
5·75		0·1434	0·1252	0·1088	0·0944	0·0822	0·0725	0·0653
6·0		0·1217	0·1036	0·0873	0·0732	0·0614	0·0522	0·0457
6·25		0·1035	0·0855	0·0695	0·0558	0·0445	·0·0359	0·0301
6·5		0·0885	0·0708	0·0552	0·0419	0·0312	0·0233	0·0183
6·75		0·0765	0·0592	0·0440	0·0313	0·0212	0·0140	0·0098
7·0		0·0671	0·0503	0·0357	0·0236	0·0142	0·0077	0·0042
7·25		0·0600	0·0438	0·0298	0·0184	0·0097	0·0039	0·0011
7·5		0·0550	0·0393	0·0260	0·0152	0·0073	0·0023	0·0001
7·75		0·0515	0·0365	0·0239	0·0139	0·0066	0·0022	0·0007
8·0		0·0493	0·0350	0·0231	0·0138	0·0072	0·0035	0·0025
8·25		0·0480	0·0344	0·0233	0·0147	0·0087	0·0056	0·0051
8·5		0·0474	0·0345	0·0241	0·0162	0·0108	0·0082	0·0081
8·75		0·0472	0·0350	0·0253	0·0180	0·0132	0·0110	0·0113
9·0		0·0472	0·0357	0·0266	0·0199	0·0156	0·0138	0·0143
9·25		0·0472	0·0363	0·0279	0·0217	0·0178	0·0163	0·0170
9·5		0·0470	0·0368	0·0289	0·0232	0·0198	0·0184	0·0192
9·75		0·0465	0·0370	0·0296	0·0243	0·0212	0·0200	0·0208
10·0		0·0458	0·0368	0·0299	0·0250	0·0222	0·0211	0·0218
10·25		0·0447	0·0362	0·0298	0·0253	0·0226	0·0216	0·0221
10·5		0·0433	0·0353	0·0293	0·0251	0·0225	0·0215	0·0219
10·75		0·0416	0·0340	0·0283	0·0244	0·0219	0·0209	0·0210
11·0		0·0396	0·0324	0·0270	0·0232	0·0209	0·0198	0·0197
11·25		0·0374	0·0305	0·0254	0·0218	0·0195	0·0183	0·0180
11·5		0·0351	0·0285	0·0236	0·0201	0·0179	0·0166	0·0161
11·75		0·0327	0·0264	0·0217	0·0183	0·0161	0·0147	0·0140
12·0		0·0303	0·0243	0·0197	0·0164	0·0142	0·0127	0·0118
12·25		0·0280	0·0223	0·0178	0·0146	0·0124	0·0108	0·0097
12·5		0·0259	0·0203	0·0160	0·0129	0·0107	0·0090	0·0078

The fourth place is unreliable at the larger values of θ.

Correction Factor

1·40	1·60	1·80	2·00	2·20	2·40	2·60	2·80	3·00
4·055	4·953	6·050	7·389	9·025	11·023	13·464	16·445	20·086
1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000	1·0000
0·9960	0·9959	0·9959	0·9958	0·9958	0·9958	0·9957	0·9957	0·9957
0·9839	0·9837	0·9835	0·9834	0·9832	0·9831	0·9831	0·9830	0·9831
0·9642	0·9638	0·9633	0·9629	0·9626	0·9624	0·9623	0·9622	0·9623
0·9371	0·9363	0·9355	0·9349	0·9344	0·9340	0·9338	0·9337	0·9338
0·9032	0·9019	0·9008	0·8999	0·8991	0·8986	0·8983	0·8982	0·8983
0·8632	0·8614	0·8598	0·8585	0·8575	0·8568	0·8565	0·8564	0·8566
0·8178	0·8154	0·8133	0·8117	0·8105	0·8097	0·8093	0·8092	0·8096
0·7678	0·7648	0·7623	0·7604	0·7589	0·7580	0·7576	0·7577	0·7582
0·7142	0·7107	0·7078	0·7055	0·7039	0·7029	0·7026	0·7029	0·7036
0·6580	0·6539	0·6506	0·6481	0·6463	0·6454	0·6452	0·6458	0·6469
0·6000	0·5955	0·5919	0·5892	0·5874	0·5866	0·5867	0·5876	0·5892
0·5414	0·5365	0·5327	0·5299	0·5282	0·5276	0·5280	0·5294	0·5316
0·4830	0·4778	0·4739	0·4711	0·4696	0·4693	0·4702	0·4722	0·4751
0·4258	0·4204	0·4164	0·4138	0·4126	0·4128	0·4142	0·4169	0·4207
0·3705	0·3650	0·3611	0·3588	0·3580	0·3587	0·3609	0·3644	0·3691
0·3178	0·3124	0·3087	0·3068	0·3066	0·3080	0·3110	0·3154	0·3211
0·2684	0·2632	0·2599	0·2585	0·2590	0·2612	0·2651	0·2705	0·2773
0·2228	0·2179	0·2152	0·2144	0·2156	0·2187	0·2236	0·2301	0·2380
0·1814	0·1769	0·1748	0·1748	0·1769	0·1809	0·1869	0·1946	0·2036
0·1444	0·1405	0·1391	0·1399	0·1429	0·1480	0·1551	0·1639	0·1741
0·1120	0·1088	0·1081	0·1098	0·1138	0·1200	0·1282	0·1382	0·1495
0·0842	0·0817	0·0819	0·0846	0·0896	0·0968	0·1061	0·1172	0·1297
0·0610	0·0593	0·0603	0·0639	0·0699	0·0783	0·0886	0·1007	0·1143
0·0421	0·0412	0·0431	0·0477	0·0547	0·0640	0·0753	0·0884	0·1029
0·0273	0·0272	0·0300	0·0355	0·0435	0·0536	0·0658	0·0798	0·0951
0·0162	0·0169	0·0206	0·0270	0·0358	0·0468	0·0597	0·0744	0·0904
0·0084	0·0100	0·0144	0·0216	0·0312	0·0428	0·0564	0·0716	0·0882
0·0036	0·0059	0·0111	0·0189	0·0291	0·0413	0·0554	0·0710	0·0879
0·0012	0·0042	0·0100	0·0184	0·0290	0·0417	0·0561	0·0719	0·0889
0·0009	0·0045	0·0108	0·0195	0·0305	0·0434	0·0579	0·0738	0·0908
0·0020	0·0061	0·0128	0·0219	0·0330	0·0459	0·0604	0·0762	0·0930
0·0043	0·0088	0·0157	0·0249	0·0360	0·0489	0·0632	0·0787	0·0951
0·0073	0·0120	0·0190	0·0282	0·0392	0·0519	0·0658	0·0808	0·0968
0·0106	0·0154	0·0225	0·0315	0·0422	0·0545	0·0679	0·0824	0·0977
0·0139	0·0188	0·0257	0·0345	0·0448	0·0565	0·0693	0·0831	0·0977
0·0170	0·0218	0·0285	0·0369	0·0466	0·0577	0·0698	0·0828	0·0966
0·0197	0·0243	0·0306	0·0385	0·0476	0·0580	0·0693	0·0815	0·0944
0·0218	0·0261	0·0320	0·0393	0·0477	0·0573	0·0678	0·0791	0·0912
0·0232	0·0272	0·0326	0·0393	0·0470	0·0557	0·0653	0·0757	0·0869
0·0240	0·0276	0·0324	0·0384	0·0453	0·0532	0·0619	0·0715	0·0818
0·0241	0·0272	0·0314	0·0367	0·0429	0·0499	0·0578	0·0665	0·0760
0·0235	0·0262	0·0298	0·0344	0·0398	0·0460	0·0531	0·0610	0·0698
0·0223	0·0245	0·0275	0·0315	0·0361	0·0416	0·0480	0·0552	0·0633
0·0206	0·0223	0·0248	0·0281	0·0321	0·0369	0·0426	0·0492	0·0568
0·0186	0·0198	0·0218	0·0244	0·0279	0·0321	0·0372	0·0433	0·0505
0·0163	0·0171	0·0186	0·0207	0·0236	0·0273	0·0320	0·0377	0·0446
0·0139	0·0143	0·0154	0·0171	0·0195	0·0228	0·0271	0·0326	0·0392
0·0114	0·0116	0·0123	0·0136	0·0156	0·0186	0·0227	0·0280	0·0345
0·0091	0·0090	0·0094	0·0105	0·0122	0·0149	0·0189	0·0241	0·0306
0·0070	0·0067	0·0069	0·0077	0·0093	0·0119	0·0158	0·0210	0·0276

The fourth place is unreliable at the larger values of θ.

It follows that the transit time correction factor can be written

$$\sum_{r=0}^{\infty} (-1)^r K_{2r}\, \theta^{2r}, \qquad\qquad \ldots\ldots(2)$$

where

$$K_{2r} = \frac{1}{(2r)!} \sum_{s=0}^{2r} (-1)^s \binom{2r}{s} J_s J_{2r-s}.$$

Using dashes to denote differentiation with respect to u_a,

$$J_n' = -(nI_a'/I_a + u_a^{-1}) J_n + u_a^{-1}$$

and, if B_n is defined as $(u_a I_a^n)^{-1} \int_0^{u_a} (I_a - I)^n\, du$,

$$B_n' = -\left(\frac{nI_a'}{I_a} + \frac{1}{u_a}\right) B_n + \frac{nI_a'}{I_a} B_{n-1}$$

and

$$\begin{aligned}
K_{2n}' &= \frac{2}{(2n)!} \sum_{s=0}^{2n} (-1)^s \binom{2n}{s} J_{2n-s} \left\{ \frac{1}{u_a} - \left(\frac{sI_a'}{I_a} + \frac{1}{u_a}\right) J_s \right\} \\
&= \frac{2}{u_a(2n)!} \left\{ \sum_{s=0}^{2n} (-1)^s \binom{2n}{s} J_{2n-s} \right\} - \left(\frac{nI_a'}{I_a} + \frac{1}{u_a}\right) \frac{2}{(2n)!} \sum_{s=0}^{2n} (-1)^s \binom{2n}{s} J_{2n-s} J_s \\
&= \frac{2}{(2n)!} B_{2n} - 2\left(\frac{nI_a'}{I_a} + \frac{1}{u_a}\right) K_{2n}.
\end{aligned}$$

It is easily seen that $J_0 = B_0 = K_0 = 1$ for all u_a and, from the circumstance that $I_a = \sqrt{u_a}(1 + \frac{1}{3} u_a + \ldots .)$, the values of B_n and K_{2n} can be deduced for small values of u_a. They are

$$B_n = \frac{2}{(n+1)(n+2)} + \frac{4}{3} \frac{n(n+7)}{(n+1)(n+2)(n+3)(n+4)} u_a + \ldots.$$

and

$$K_{2n} = \frac{8(2n+3)}{(2n+4)!} + \frac{128}{3} \frac{n(2n+5)}{(2n+6)!} u_a + \ldots..$$

When I_a'/I_a had been found, the above differential equations for B_n and K_{2n} were solved (given the initial values) numerically to furnish, first B_1, B_2,.... in succession, and secondly, using the values of the B_{2n}'s, the values of K_{2n}. When the K_{2n}'s are known, the transit time correction factor may be calculated from expression (2).

It only remains to explain how

$$\frac{I_a'}{I_a} = \frac{\frac{1}{2} u_a^{-\frac{1}{2}} \exp u_a}{\displaystyle\int_0^{\sqrt{u_a}} \exp(x^2)\, dx}$$

was tabulated. Of the three functions occurring in this expression, available tabulation is the least accurate for the denominator. For a time, use was made of the table of Terrill and Sweeney (1944) which gives I_a to 6 places of decimals at intervals of 0·01 in the upper limit, $\sqrt{u_a}$. However, after $(nI_a'/I_a + u_a^{-1})$ had been obtained for $n = 1$, 2 and 3, it became evident that, for the accuracy and range required in the tabulation of the transit time correction factor, use of a 6-place table was insufficient. A new table was therefore made according to the following scheme:

(i) $\quad u_a^{-\frac{1}{2}} \displaystyle\int_0^{\sqrt{u_a}} \exp(x^2)\, dx = \left(1 + \frac{u_a^2}{10} + \ldots . \right) + \left(\frac{u_a}{3} + \frac{u_a^3}{42} + \ldots . \right).$

The two terms were evaluated separately at intervals of 0·2. Their difference, when multiplied by $2(u_a/\pi)^{\frac{1}{2}}$, should equal $\mathrm{erf}(\sqrt{u_a})$ and this was verified for each value of u_a.

(ii) The sum of the two terms was divided into $\exp u_a$, giving $2u_a I_a'/I_a$ at 0·2 intervals to 11 places of decimals. This table was differenced, and a table prepared from it at the interval chosen for the numerical integration of the B and K differential equations.

(iii) The entries in this last table were differenced and divided by $2u_a$ to give I_a'/I_a.

§4. DISCUSSION OF ACCURACY

In the whole computation there were five separate processes:

(i) evaluation of I_a'/I_a and $(nI_a'/I_a + u_a^{-1})$;
(ii) solution of the differential equations for B_n;
(iii) solution of the differential equations for K_{2n};

(iv) substitution in the power series $\sum_{r=0}^{\infty} (-1)^r K_{2r} \theta^{2r}$ at key points;

(v) interpolation to give the transit time correction factor at intermediate points.

Operation (i) has been explained above and should be free from error.

It will be seen from the nature of the B and K differential equations that the small inevitable errors due to rounding do not grow as u_a increases but that errors in a B-table are passed on, with the same sign and a reduction in magnitude, to the next B-table and to the K-table. It is believed that none of the K's will be wrong by more than a unit or two in the last place and that the errors which do occur will tend to have the same sign for consecutive K's at the same value of u_a and therefore affect the transit time correction factor in opposite directions. The number of decimal places in each K-table is shown in Table 3, together with the effect of a unit error in the last place of each one.

Table 3

Effect on transit time correction factor of an error $+1$ in last place.

$2n$	Decimals in K_{2n}	at $\theta=8$ radians	at $\theta=10$ radians	at $\theta=12\frac{1}{2}$ radians
2	7	−0·0000 06	−0·0000 10	−0·0000 16
4	9	+0·0000 04	+0·0000 10	+0·0000 24
6	11	−0·0000 03	−0·0000 10	−0·0000 38
8	13	+0·0000 02	+0·0000 10	+0·0000 60
10	15	−0·0000 01	−0·0000 10	−0·0000 93
12	17	+0·0000 01	+0·0000 10	+0·0001 46
14	19	0·0000 00	−0·0000 10	−0·0002 27
16	22	0·0000 00	+0·0000 01	+0·0000 36
18	24	0·0000 00	−0·0000 01	−0·0000 56
20	26	0·0000 00	+0·0000 01	+0·0000 87
22	28	0·0000 00	−0·0000 01	−0·0001 35
24	31	0·0000 00	0·0000 00	+0·0000 21
26	33	0·0000 00	0·0000 00	−0·0000 33
28	35	0·0000 00	0·0000 00	+0·0000 52
30	37	0·0000 00	0·0000 00	−0·0000 81

Errors in the fourth decimal place of the factor arising from this source should therefore not amount to more than the occasional turning of a figure up to 8 radians, 1 at 10 radians and 4 at the extreme range of the table.

Errors in interpolation between key points were reduced by differencing both rows and columns of the table and checking doubtful values by substitution in the power series.

If it were desired to extend Table 2 to larger transit angles by this power series method, more significant figures would be required in the tables of the coefficients K_{2n}. This being so, the alternative method of using an asymptotic series would probably be found easier if any such extension were made.

§5. THE ZERO POINT

The small values of the factor at $u_a = 1\cdot2$, $\theta = 7\cdot5$ and at nearby points in the table suggest that there might be a point in that neighbourhood for which the factor is zero. As a first step in a detailed investigation of that neighbourhood, the integrals $u_a^{-1} \int_0^{u_a} \cos(I\theta/I_a)\,du$ and $u_a^{-1} \int_0^{u_a} \sin(I\theta/I_a)\,du$ were evaluated separately by direct computation for $u_a = 1\cdot2$, $\theta = 7\cdot5$. The sum of their squares was in excellent agreement with the value of the transit time correction factor given in the table for this same point, a useful check on the whole process by which the tabular values were obtained. Then, keeping the integrands unchanged but varying the upper limit of integration a little (which corresponds to giving θ and u_a small increments such that θ/I_a remains unchanged), a point was found at which the cosine integral vanished and another point at which the sine integral vanished. A slightly different value of θ/I_a was next assigned and a second null point was found similarly for each integral. A first approximation to the coordinates (θ, u_a) of the point of intersection of the nodal lines of the cosine and sine integrals was obtained by supposing that each of the nodal lines was straight between the pair of null points already found on it. This approximation gave a value of θ/I_a which was used as before to find a null point for each of the integrals; these two null points were nearer together and led to a better approximation for the point of intersection. Repetition of this process showed that the transit time correction factor vanishes when $u_a = 1\cdot246$ and $\theta = 7\cdot498$ radians. Near this point, the value of the factor is approximately

$$0\cdot0128(\theta - 7\cdot498)^2 + 0\cdot0126(\theta - 7\cdot498)(u_a - 1\cdot246) + 0\cdot0363(u_a - 1\cdot246)^2.$$

REFERENCES

KOMPFNER, R., HATTON, J., SCHNEIDER, E. E., and DRESEL, L. A. G., 1946, *J. Instn. Elect. Engrs.*, **93**, Pt. III A, 1436.

RAMO, S., 1939, *Proc. Inst. Radio Engrs.*, **27**, 584.

SHOCKLEY, W., 1938, *J. Appl. Phys.*, **9**, 635.

SPENKE, E., 1937, *Wissen. Veroff. Siemens-Werken*, **16**, 127.

TERRILL, H. M , and SWEENEY, L., 1944, *J. Franklin Inst.*, **238**, 220 (cf. *Math. Tables and Aids to Comp.*, **1**, 1944, pp. 322–23).

The Mode of Fracture at the Neck of a Tensile Specimen

By W. I. PUMPHREY

School of Industrial Metallurgy, University of Birmingham

ABSTRACT. Bridgman has shown that in a cylindrical specimen which is deforming plastically under longitudinal tension the tensile stress at the neck of the specimen is greatest on the longitudinal axis and least at the periphery of the specimen. Fracture therefore commences at the centre of the specimen and extends from the centre to the periphery. " Double-cup " fractures encountered during the tensile testing of a number of alloys at elevated temperatures have been accounted for in this way and an examination has been made of the metallurgical factors which govern the occurrence of this type of fracture.

---

§ 1. INTRODUCTION

DURING an investigation into the mechanical properties of alloys in a number of binary aluminium alloy systems at temperatures between 400° c. and the temperature of incipient melting, the occurrence of an unusual type of fracture at the neck of the tensile specimen was observed in a considerable number of instances. The specimens employed for the high temperature tensile testing were of the form shown in Figure 1, and after the majority
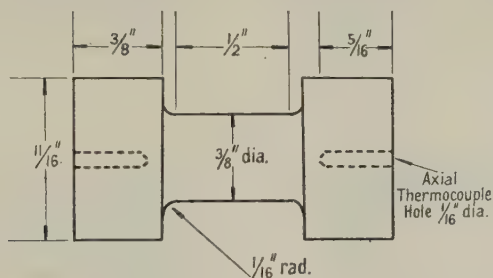
Figure 1. Design of test piece used in high temperature tensile tests.

of the tensile tests each of the two surfaces of fracture in the broken specimens contained a central cavity of some depth. The two portions of a tensile specimen of an aluminium–copper alloy containing 0·5% copper which fractured in this way at a testing temperature of 506° c. are shown in Figure 2 (Plate I).

The " double-cup " type of fracture was thought to be due to the initiation of fracture at the centre of the specimen, followed by considerable elongation of the resulting internal cavity at the high temperature of testing and the ultimate fracture of the " tube " of metal so formed in such a manner as to give a central cavity in each of the two surfaces of fracture in the broken specimen. The fact that, under certain conditions, fracture can commence on the central axis of a necked specimen during tensile testing has been known for some years, and has been utilized to provide an explanation of the " cup and cone " fractures which occur in many brasses and steels tensile tested at room temperatures. An early reference to the phenomenon was made by Ludwik (1927), and a mathematical analysis of the stress distribution at the neck of a tensile specimen was later

42–2

advanced by Bridgman (1944) who showed that in a cylindrical specimen deforming plastically under longitudinal tension the tensile stress at the neck of the specimen is greatest on the longitudinal axis and least at the periphery. There appears to have been no previous observation of the double-cup type of fracture and it was therefore felt that an examination of the metallurgical factors which govern the occurrence of this form of fracture would be of value.

§ 2. EXPERIMENTAL WORK

It was thought to be of importance to determine firstly whether the occurrence of fracture on the axis of a tensile specimen prior to the complete fracture of the specimen is initiated or aggravated by the presence of central unsoundness in the specimen due to contraction cavities or gas porosity in the material. It was also considered to be of importance to determine at what stage fracture commences on the central axis during the process of elongation. Five specimens of an aluminium–iron alloy containing 0·5% iron were therefore extended by different amounts varying from 21% to 65% at a temperature of 632° c., sectioned longitudinally and examined under the microscope. Four of the five specimens were observed to contain one or more cavities along their central axes after elongation, the fifth being quite sound. The soundness of this latter specimen, which was elongated by 58% before sectioning, was somewhat anomalous since the specimen which was extended by only 21% was found to contain numerous central cavities. This result suggested that the occurrence of fracture on the axes of the tensile specimens might be in some way associated with central unsoundness in the specimens before testing. In order to examine the question further, three specimens of an aluminium–iron alloy containing 1% iron were extended to the point of fracture in the tensometer at a testing temperature of 630° c., the specimens being removed from the machine and radiographed at regular intervals during the process of extension. The three specimens were also radiographed before testing. There was evidence of porosity in one of the specimens before testing although the remaining two appeared to be quite sound. After an extension of 20% at the temperature of testing a number of central cracks and cavities appeared in all three specimens. After further extension, the radiographic examination revealed that one of the cracks in each specimen was considerably enlarged—see Figure 3 (Plate I)—and this crack was eventually observed to develop into an elongated central cavity which finally gave rise to a double-cup type of fracture.

The initial observations appeared to indicate that there was some central unsoundness in the cast bars from which the specimens were machined and an attempt was therefore made to produce a bar free from such unsoundness. To this end, an ingot of an aluminium–iron alloy containing 0·5% iron was cast, the metal being thoroughly degassed before being poured into the mould. The ingot so produced was forged to effect a reduction in cross-sectional area of some 75% and six tensile specimens were machined from the forged bar to the form shown in Figure 1. Radiographic examination of the specimens before tensile testing indicated that all were free from central unsoundness. The specimens were extended at a temperature of 630° c. and were radiographed at intervals during the process of extension. After all the specimens had been reduced in cross-sectional area by some 90% only one was observed to contain a central cavity. Each of the six specimens fractured after a reduction in area of approximately 98% and in five of the six specimens the final fractures were of a pronounced
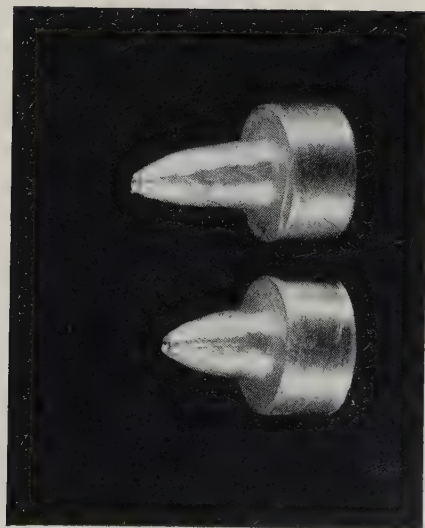
Figure 4. Section through fracture in an aluminium–copper alloy containing 2% copper. Etched in a 0·5 aqueous solution of hydrofluoric acid. Magnification ×10.
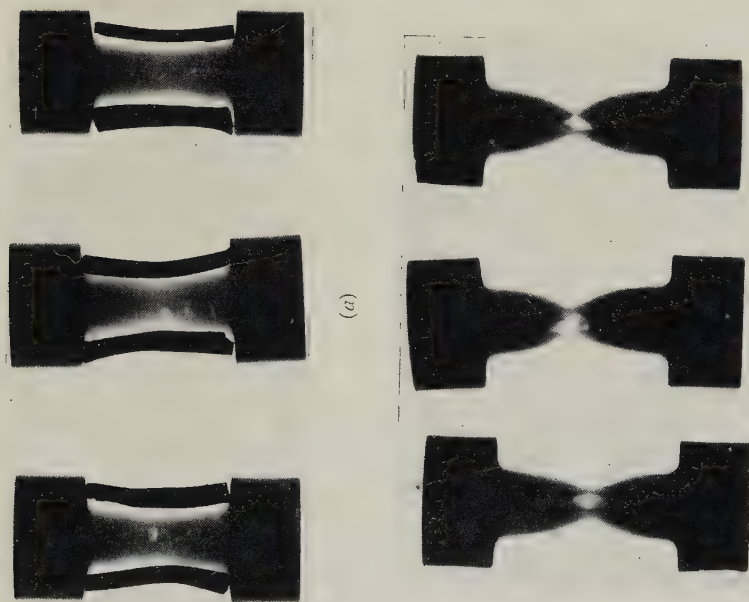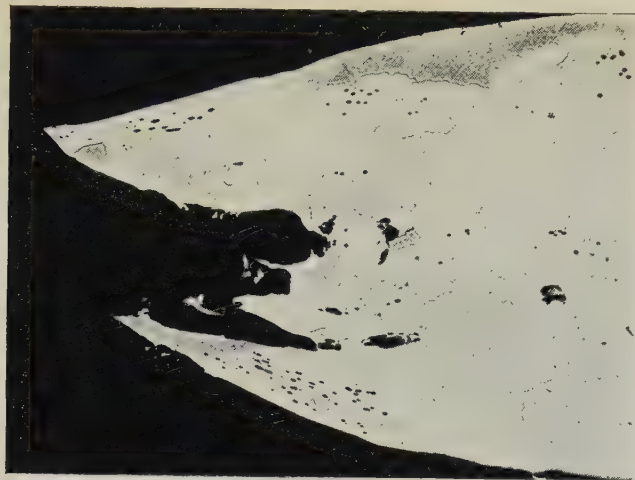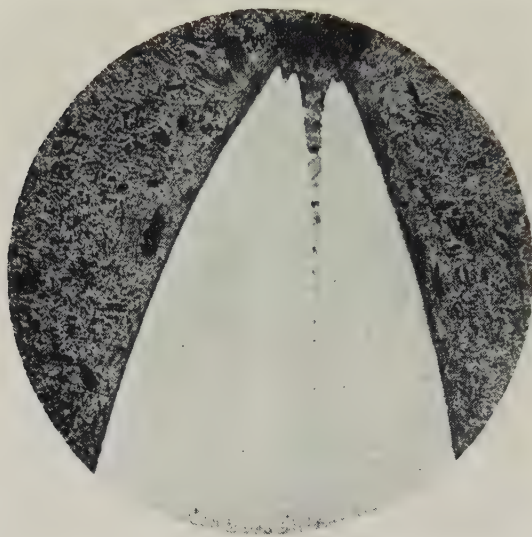

(a)

(b)

Figure 3. Radiographs of test pieces of an aluminium–iron alloy containing 1% iron showing the development of central cavities. (a) After 20% extension at 630° C. (b) After 60% extension at 630° C.

Plate I.

Figure 2. A fractured test-piece of an aluminium–copper alloy containing 0·5% copper showing a "double-cup" fracture. Actual size.

Figure 5. Section through fracture in a wrought mild steel specimen tested at 973° c. Unetched. Magnification ×15.
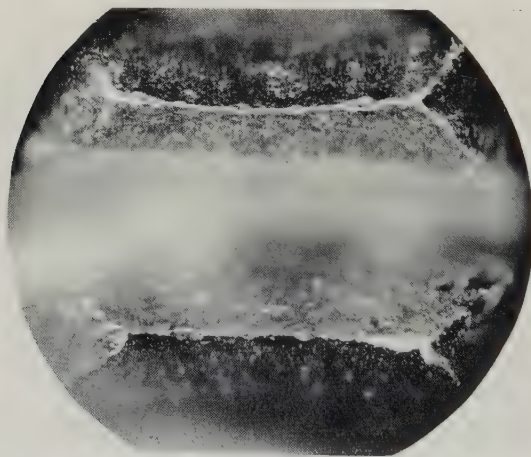
Figure 6. Appearance of fracture in both portions of a lead strip test piece tested at room temperature. Magnification ×12.

Plate II.

double-cup form. Inside the cavity in both surfaces of fracture in the five speci-
mens which gave a double-cup type of break there was a number of holes in a
type of "honey-comb" formation : the reason for this latter effect is not yet clear.

The results of the initial experiments would seem to indicate that fracture
can commence on the axis of a specimen during tensile testing in the absence of
any central unsoundness in the specimen but that its occurrence may be facilitated
by the presence of such unsoundness.

It should here be remarked, however, that although radiographic examination
may fail to reveal the presence of any unsoundness in a specimen before tensile
testing it is not entirely safe to assume that the specimen is completely free from any
minute central cracks or flaws since a radiographic method of examination may
not be sufficiently sensitive for the detection of very minute flaws.

From the work of Bridgman it would seem that in a cylindrical specimen
subjected to longitudinal tension the development of an axial tension at the centre
greater than that at the periphery, and hence the initiation of fracture on the
central axis, can occur at the neck of the specimen provided that the material
possesses sufficient ductility for necking to occur. The initiation of fracture
on the central axis and the development of a double-cup type of break would
also seem to be dependent upon there being no unduly excessive strain-hardening
of the material for any given extension. In order to examine the validity of these
deductions specimens were machined to the design shown in Figure 1 from
a cast bar of lead and were tensile tested at a number of different temperatures
between room temperature and the temperature of incipient melting. At all
temperatures of testing below that at which melting of the specimen began the
test pieces elongated to a considerable extent before fracture occurred. A
pronounced central cavity was observed in both surfaces of fracture in all the
specimens tested at temperatures below about 250° c. At temperatures above
250° c. the tensile specimens elongated by at least 100% and necked down to a
point before fracture occurred : no double-cup type of fracture was observed at
such temperatures of testing. Tensile tests carried out at room temperatures
and at elevated temperatures on specimens of magnesium and beryllium, which
have a hexagonal crystal lattice and are not inherently ductile metals even at
elevated temperatures, confirmed that, in the absence of adequate ductility,
fracture is not initiated on the central axis of the tensile specimen, and that, in
consequence, the double-cup type of fracture does not occur.

In a cylindrical specimen which is deforming plastically under longitudinal
tension the material on the longitudinal axis at the neck of the specimen is stressed
in three directions at right angles and consequently fractures in a brittle manner
when the axial tension has risen sufficiently high. The two surfaces of the central
fracture therefore have all the characteristics of fracture in a non-ductile material.

This is illustrated by Figure 4 (Plate I) which shows a section through the
central cup in an aluminium–copper alloy containing 2% copper tested at a tem-
perature of 522° c. As will be seen, the crystals at the periphery of the test piece
are considerably elongated, indicating that in these regions the metal deformed
plastically ; at the base and at the sides of the cup the crystals are much
more equiaxial, indicating that the material on the central axis did not elongate
under the applied stress but fractured in a brittle, intercrystalline manner at an
early stage in the elongation of the test piece. The brittle fracture subsequently
extends towards the outside of the specimen under the applied tension until

the geometry of the specimen is so modified that fracture of the outer layers occurs by shear. Whether the specimen will extend sufficiently before the completion of the shearing mechanism for a double-cup type of fracture to occur is dependent upon the ductility of the material and to some small extent upon its strain-hardening characteristics. Steels and brasses are, in general, more ductile than magnesium or beryllium but less so than lead, and in tensile tests of such metals at room temperature the occurrence of brittle fracture on the central axis is followed almost immediately by shear fracture of the outer layers. The shearing mechanism is completed without any further appreciable extension of the specimen, so giving the characteristic cup and cone fractures usually observed in cylindrical specimens of steel and brass fractured under longitudinal tension at normal temperatures. By tensile testing at suitably elevated temperatures, the ductility of steel and brass is increased, strain-hardening is instantaneously relieved, and the specimen extends by a considerable extent before the final shear fracture of the outer layers, so giving a final fracture of double-cup form. A section through one half of a fractured tensile specimen of wrought mild steel tested at a temperature of 973° c. is shown in Figure 5 (Plate II). The other half of the specimen was exactly similar in appearance. Figure 5 also illustrates an extension of the double-cup phenomenon which has been observed with a variety of metals tensile tested at elevated temperatures, namely that after the commencement of fracture at the centre of a cylindrical specimen, fracture can subsequently occur in an exactly similar manner at the centre of the wall of the "tube" formed as a result of the longitudinal extension of the initial central fracture. The narrow band of material extending downwards from the central cavity in the specimen shown in Figure 5 was examined microscopically and a number of sulphur prints taken of the specimen as a whole. This examination indicated that the central band was almost entirely ferritic and contained an appreciable amount of manganese sulphide and other sulphur bearing compounds. In this instance the presence of a central band of impurity in the cylindrical specimen may have influenced the initiation of fracture on the axis.

It might be expected that the shape of the specimen as well as the ductility of the material would have a considerable effect on the tendency to necking, on the stress distribution at the neck, and hence on the development of the double-cup type of fracture. In order to determine the effect of specimen shape tensile tests were carried out on test pieces cut from material in sheet form and on cylindrical specimens having a considerably elongated gauge length.

Test pieces cut from lead sheet 0·080 in. in thickness which were tensile tested at room temperature gave, in general, a reduction of area of approximately 100% and fractured with a line fracture. The appearance of the fracture in both portions of one specimen is shown in Figure 6 (Plate II). As can be seen, there is a series of cavities along the line of fracture and it would seem that in this instance also fracture commenced at the centre of the specimen.

Tensile tests were carried out at temperatures of 488° c. and 621° c. respectively on two cylindrical specimens of an aluminium–iron alloy containing 1% iron, the gauge length of each specimen being four times that of the specimens machined to the design shown in Figure 1. The specimen tested at the lower temperature gave a reduction in area of 98% before fracture occurred and had no apparent tendency to a double-cup type of fracture. The other specimen gave a reduction in area of 85% and there was a marked central cavity in both surfaces of fracture in the broken test piece.

§3. DISCUSSION AND CONCLUSIONS

Bridgman has shown that in a cylindrical specimen which is deforming plastically under longitudinal tension the tensile stress at the neck is greatest on the longitudinal axis and least at the periphery of the specimen. The material on the longitudinal axis at the neck of the specimen is stressed in three directions at right angles and consequently fractures in a brittle manner when the axial stress has risen sufficiently high. In a metal tested at a temperature such that the ductility is low and considerable strain-hardening can occur; the occurrence of brittle fracture on the central axis is followed almost immediately by shearing of the outer layers giving the characteristic cup and cone type of fracture. During the course of the present work it has been found that if a metal is tensile tested at a temperature such that the ductility is high and strain-hardening is instantaneously relieved, there is a tendency for the specimen to elongate considerably after fracture has begun on the central axis. Tensile testing under such conditions may cause the specimen to break finally with a double-cup type of fracture.

It is difficult to determine the exact stage during the elongation at which fracture begins on the central axis of a cylindrical specimen under longitudinal tension. If the occurrence of central fracture is due to the Bridgman stress effect then fracture at such points cannot be considered to occur until the specimen is deforming plastically, that is, until the material has been stressed beyond its elastic limit. If, however, fracture on the central axis is initiated by central unsoundness, then, due to the concentration of stress around an internal cavity in a tension member, such fracture can be imagined to occur at a stress lower than the elastic limit of the material and the extension of the fracture cavities under tension may, in fact, contribute to neck formation. During the course of the present investigation the double-cup type of fracture has been observed in some hundreds of specimens and it is unlikely that every specimen in which this type of fracture occurs contains some central unsoundness. It would seem rather that the occurrence of central fracture is not dependent upon the specimen containing any central unsoundness but is facilitated by the presence of such unsoundness. The initiation of fracture on the central axis of a tensile specimen thus appears to be principally dependent upon the metal having sufficient ductility for some necking to take place. The final form of fracture, whether cup and cone or double-cup, is determined by the ductility of the metal, and, to some extent, by the amount of strain-hardening which occurs in the metal prior to complete fracture, the greater the ductility and the lower the tendency to strain-hardening, the greater the tendency for the occurrence of a double-cup type of break. The complete absence of any tendency to strain-hardening, however, does not seem essential for fracture to commence at the centre of the specimen.

REFERENCES

BRIDGMAN, P. W., 1944, *Trans. Amer. Soc. Metals*, **32**, 553.
LUDWIK, P., 1927. *Z. Ver. dtsch. Ing.*, **71**, 1532.

The Magnetostriction of Anisotropic Permanent Magnet Alloys

By M. McCAIG

Permanent Magnet Association, Central Research Laboratory, Sheffield

MS. received 26th May 1949

ABSTRACT. Blocks of permanent magnet alloys of the system Fe Ni Al Co Cu have been prepared with columnar crystals. The magnetostriction has been measured in various directions before and after heat treatment in a magnetic field. The results agree well with those predicted theoretically, and so confirm the postulated crystal structure.

§1. INTRODUCTION

A NUMBER of ferromagnetic alloys can be produced which possess magnetic anisotropy, i.e. have in different directions different values of magnetic characteristics such as remanence, coercive force, permeability and magnetostriction. This paper is concerned with anisotropy due to two causes. Firstly, there is anisotropy due to a preferred crystal orientation. A well-known example of anisotropy due to this cause is provided by cold-rolled silicon–steel strip. Secondly, there is anisotropy due to a preferred orientation of the magnetic domains. This preferred domain orientation is produced by cooling certain alloys in a magnetic field. Alloys which respond to this treatment include certain iron–nickel alloys as well as those more complicated ones used for the manufacture of anisotropic permanent magnets. The above two causes of anisotropy may exist either separately or together.

Hoselitz and McCaig (1949) have published an account of an investigation of the magnetostriction of an alloy in which a preferred domain orientation was produced. The method of measuring magnetostriction, using a roller and optical lever, which was described in that paper, has again been used in the present work.

The alloys investigated range from Alcomax II with a nominal composition Ni $11 \cdot 0^{\circ}{}_{0}$, Co $24 \cdot 0^{\circ}{}_{0}$, Cu $6 \cdot 0^{\circ}{}_{0}$, Al $8 \cdot 0^{\circ}{}_{0}$, balance Fe, to Hycomax with a nominal composition Ni $21 \cdot 0^{\circ}{}_{0}$, Co $20 \cdot 0^{\circ}{}_{0}$, Cu $2 \cdot 0^{\circ}{}_{0}$, Al $9 \cdot 0^{\circ}{}_{0}$, balance Fe, and including Alcomax III which contains a few per cent of niobium (columbium). In an isotropic state these alloys have a positive longitudinal magnetostriction of the order 1 to 2×10^{-5}.

It was first suggested by Dr. K. Hoselitz of the Permanent Magnet Association that if the existing anisotropic permanent magnet alloys could be obtained with columnar crystals, higher values of the remanence, coercive force and energy would result. A material with columnar crystals is polycrystalline, but all the crystals have one axis parallel to a given direction. A considerable improvement in the desired magnetic properties has in fact been obtained by producing an alloy with columnar crystals and then cooling in the usual manner in a magnetic field parallel to the common crystal axis.

Several alloys within the range specified above have been obtained with columnar crystals, and the magnetostriction of these alloys has been measured. Although the figures quoted below refer to one particular alloy, the results are qualitatively similar for the others investigated.

The specimens used were approximately cubical blocks of 2·5 to 3 cm. edge. These blocks were cut in such a way, that, as could be seen by visual inspection, the majority of the crystal boundaries were parallel to one edge of the block. Thus if the Figure represents one of the blocks, the majority of the crystals were elongated
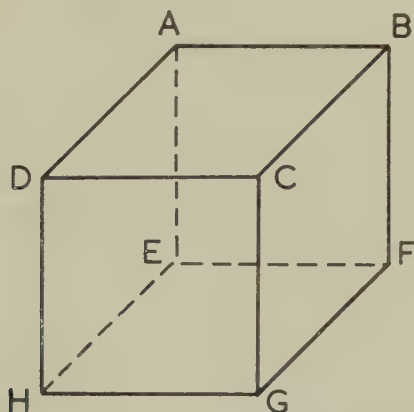
and lay parallel to one edge, say AB. Assume provisionally that the crystals are body-centred cubic and that they have a [100] direction, that is a cube edge, parallel to AB, so that we can refer to AB as the common crystal axis. The remaining cube edges of the crystals must be in the plane parallel to ADHE, but the orientation of the crystal axes in this plane is expected to be random. Finally it is assumed that the crystals are iron-like, and that the [100] cube edge direction is the easy direction of magnetization. The assumption of body-centred-cubic (iron-like) crystals is plausible, because this structure has been found by x-ray analysis in alloys of not very different composition (Oliver and Goldschmidt 1946), but the accuracy or inaccuracy of the above picture of the crystal orientation rests on the degree of success with which it enables the observed measurements of magnetostriction to be predicted*.

§2. THEORY AND EXPERIMENT

An approximate theory of magnetostriction (Becker and Döring 1939) which has been found to work reasonably well for iron-like substances gives for the saturation magnetostriction of a single crystal

$$\left(\frac{dl}{l}\right)_{\alpha_i\beta_i} = \frac{3}{2}\lambda_{100}(\alpha_1^2\beta_1^2 + \alpha_2^2\beta_2^2 + \alpha_3^2\beta_3^2 - \tfrac{1}{3})$$
$$+ 3\lambda_{111}(\alpha_1\alpha_2\beta_1\beta_2 + \alpha_2\alpha_3\beta_2\beta_3 + \alpha_3\alpha_1\beta_3\beta_1). \qquad \cdots\cdots(1)$$

In this equation $(dl/l)_{\alpha_i\beta_i}$ is the fractional change of length in a direction whose direction cosines are $\beta_1\beta_2\beta_3$ when the crystal is magnetized to saturation in a direction whose direction cosines are $\alpha_1\alpha_2\alpha_3$, the direction cosines being measured from the cube edges of the crystal. λ_{100} and λ_{111} are the saturation magnetostriction in the [100] and [111] directions respectively. In order to apply equation (1) to a polycrystalline material each of the quartic terms, such as $\alpha_1^2\beta_1^2$,

---

* *Note added in proof.* Mr. H. J. Goldschmidt has kindly examined one of the specimens by x-ray diffraction, and has confirmed a body-centred cubic structure with a [100] direction parallel to AB.

must be averaged over all the individual crystals. For a completely random crystal orientation the well-known result

$$\left(\frac{dl}{l}\right)_{\text{long}} = \frac{2\lambda_{100} + 3\lambda_{111}}{5} \qquad \cdots\cdots (2)$$

and

$$\left(\frac{dl}{l}\right)_{\text{trans}} = -\frac{2\lambda_{100} + 3\lambda_{111}}{10} \qquad \cdots\cdots (3)$$

is obtained.

In Table 1 columns (1) and (2) give respectively the direction in which the block was magnetized and the direction in which the change in length was measured. Column (3) gives dl/l in terms of λ_{100} and λ_{111} as deduced from equation (1). Some of the values given in this column can be checked by inspection. Thus when the directions of magnetization and measurement are both parallel to AB, $\alpha_1 = \beta_1 = 1$ and $\alpha_2 = \alpha_3 = \beta_2 = \beta_3 = 0$ for all the crystals, and when these values are inserted in equation (1) the magnetostriction is seen at once to be λ_{100}. For some directions the values of the α's and β's vary from crystal to crystal, and the mean value of the quartic terms is obtained by a simple integration. The experimental values in column (4) meas. are means taken over the two blocks and usually several faces. It was found that the first four cases were best fitted

Table 1.

Case	(1)	(2)	(3)	(4) calc.	meas.
1	AB	AB	λ_{100}	35·2	34·2
2	AB	AD or AE	$-\lambda_{100}/2$	−17·6	−17·3
3	AD	AD	$\frac{5}{8}\lambda_{100} + \frac{3}{8}\lambda_{111}$	25·2	25·2
4	AD	AB	$-\lambda_{100}/2$	−17·6	−18·3
5	AD	AE	$-\frac{1}{8}\lambda_{100} - \frac{3}{8}\lambda_{111}$	−7·2	−7·5
6	Sample with	long.	$\dfrac{2\lambda_{100} + 3\lambda_{111}}{5}$	19·3	20
7	random crystals	trans.	$-\dfrac{2\lambda_{100} + 3\lambda_{111}}{10}$	−9·6	−8

(1) Direction of magnetization; (2) Direction of measurement; (3) Saturation magnetostriction calculated in terms of λ_{100} and λ_{111}; (4) Numerical value of saturation magnetostriction $\times 10^6$.

by the values $\lambda_{100} = 35\cdot2 \times 10^{-6}$, $\lambda_{111} = 8\cdot6 \times 10^{-6}$. Since λ_{111} never appears with a larger coefficient than $3/8$, its value is less accurate than that of λ_{100}. The test of the theory and with it the validity of the postulate on crystal orientation lies in the fifth row of figures where the low calculated value of $-7\cdot2 \times 10^{-6}$ agrees well with the experimental value of $-7\cdot5 \times 10^{-6}$. These figures apply when the direction of magnetization and the direction of measurement are perpendicular both to the common crystal axis, and to each other. If there were no crystal orientation the longitudinal magnetostriction would have the same value in all directions and the transverse magnetostriction would have half this value.

The blocks were next subjected to the treatment known in industry as magnetic hardening. That is, they were cooled from a high temperature of about 1,300° c. in a magnetic field. Block 1 was cooled with the magnetic field parallel to the edge AB, whereas block 2 was cooled with the edge AD parallel to the field.

Thus in block 1 the common crystal axis and the preferred direction of domain orientation coincided, but in block 2 they were perpendicular. Magnetic hardening increases the remanence of a permanent magnet alloy. It was shown by McCaig and Hoselitz (1949) that with random crystal orientation we should expect in the preferred direction that $J_r = 0.835 J_s$ where J_r and J_s are the intensities of magnetization at remanence and saturation respectively. This result was obtained by assuming that at the remanence point all the domains are magnetized along the nearest cube edge to the preferred direction. The same assumptions would give, for block 1, $J_r = J_s$ and for block 2, $J_r = 0.9 J_s$. The experimental values were $J_r = 0.94 J_s$ and $J_r = 0.93 J_s$ respectively.

The calculation of the magnetostriction of the blocks was now complicated by the fact that even in the demagnetized state the domain magnetization was not randomly distributed about the crystal axis. In consequence equation (1) must be replaced by

$$\left(\frac{dl}{l}\right)_{\alpha_i \beta_i} = \frac{3}{2} \lambda_{100}(\alpha_1^2\beta_1^2 + \alpha_2^2\beta_2^2 + \alpha_3^2\beta_3^2 - \alpha_1'^2\beta_1'^2 - \alpha_2'^2\beta_2'^2 - \alpha_3'^2\beta_3'^2)$$

$$+ 3\lambda_{111}(\alpha_1\alpha_2\beta_1\beta_2 + \alpha_2\alpha_3\beta_2\beta_3 + \alpha_3\alpha_1\beta_3\beta_1 - \alpha_1'\alpha_2'\beta_1'\beta_2' - \alpha_2'\alpha_3'\beta_2'\beta_3' - \alpha_3'\alpha_1'\beta_3'\beta_1').$$

$$\cdots \cdots (4)$$

Here the primes indicate the initial demagnetized state. The method of calculation is much the same as before, although the actual labour was a little more involved. Theory and experiment are compared in Table 2, the arrangement of

Table 2.

Case	Block	(1)	(2)	(3)	(4) calc.	meas.
1	1	AB	AB	—	0·0	−0·3
2	1	AB	AD or AE	—	0·0	0·0
3	1	AD	AD	$\frac{9}{8}\lambda_{100} + \frac{3}{8}\lambda_{111}$	43·9	44·0
4	1	AD	AB	$-\frac{3}{2}\lambda_{100}$	−54·6	−61·0
5	1	AD	AE	$\frac{3}{8}\lambda_{100} - \frac{3}{8}\lambda_{111}$	10·7	28·3
6	2	AD	AD	$-0.102\lambda_{100} + \frac{3}{8}\lambda_{111}$	−0·8	3·5
7	2	AD	AB	—	0·0	−3·0
8	2	AD	AE	$0.102\lambda_{100} - \frac{3}{8}\lambda_{111}$	0·8	3·5
9	2	AB	AB	$\frac{3}{2}\lambda_{100}$	54·6	54·0
10	2	AB	AD	$-1.23\lambda_{100}$	−44·8	−39·6
11	2	AB	AE	$-0.273\lambda_{100}$	−9·9	−10·0
12	2	AE	AE	$0.85\lambda_{100} + \frac{3}{8}\lambda_{111}$	33·9	40·1
13	2	AE	AB	—	0·0	5·4
14	2	AE	AD	$-0.85\lambda_{100} - \frac{3}{8}\lambda_{111}$	−33·9	−33·2

(1) Direction of magnetization; (2) Direction of measurement;
(3) Saturation magnetostriction calculated in terms of λ_{100} and λ_{111};
(4) Numerical value of saturation magnetostriction $\times 10^6$.

which is similar to Table 1. Cases 4, 7, 9, 10 and 11 involved λ_{100} but not λ_{111}. The values of λ_{100} calculated for these cases vary from 32 to 41×10^{-6}, the mean being 36.4×10^{-6}. Using this value of λ_{100}, λ_{111} can be calculated from cases 3, 5, 6, 8, 12 and 14. The value obtained from case 5 is -39×10^{-6}, but the remaining values lie between 0 and $+24 \times 10^{-6}$. The values are again inaccurate because o

the small coefficient of λ_{111}, and a better picture of the agreement between theory and experiment is obtained by calculating the magnetostriction in each case from the mean values of λ_{100} and λ_{111}. The mean values used were $\lambda_{100} = +36\cdot4 \times 10^{-6}$ and $\lambda_{111} = +7\cdot8 \times 10^{-6}$. In obtaining these figures case 5 was ignored, since it was suspected that the experimental value was in error. Not only did this case give the large negative value of λ_{111}, but in conjunction with cases 3 and 4 it would require a large volume magnetostriction. A separate experiment showed that the volume magnetostriction was in fact very small. Thus one of the experimental values belonging to case 3, 4 or 5 must be in error, and it is reasonable to assume that it was that of case 5. The experiment could not be repeated since, by the time the discrepancy was noticed the specimen had undergone further heat treatment.

With this method of comparison some of the calculated and theoretical values of the magnetostriction must agree. That a pair of values of the constants λ_{100} and λ_{111} can be found which predict the variation of the magnetostriction with direction so accurately is quite remarkable. This variation with direction is quite different from that found when crystal orientation is absent. Even the much criticized case 5 is qualitatively correct. This is the only case met in which the transverse magnetostriction is found to have a substantial positive value, and the theory does predict a positive value. Without some kind of crystal orientation it is inconceivable that the transverse magnetostriction should have the same sign as the longitudinal. The very large negative magnetostriction found in case 4 is also predicted, although it is equally inconceivable without crystal orientation.

There are many factors which might have prevented an exact agreement between theory and experiment. Normal experimental errors are likely to be of the order 5 to 10%, but greater errors may be produced if there are small cracks in the specimens. Some individual crystals certainly do not conform to the general orientation postulated, so that the calculated values of magneto-striction are not exact. The formula used is also an approximate one ignoring higher terms. In view of these considerations the agreement obtained is very satisfactory, and confirms the postulated crystal structure of the blocks.

After tempering qualitatively similar results are obtained, but the quantitative agreement cannot be expected to be quite so good, since it is likely that the domains will not be so parallel to the crystal axes.

REFERENCES

BECKER, R., and DORING, W., 1939, *Ferromagnetismus* (Berlin : Springer), p. 280.
HOSELITZ, K., and McCAIG, M., 1949, *Proc. Phys. Soc.* B, **62**, 163.
OLIVER, D. A., and GOLDSCHMIDT, H. J., 1946, *E.R.A. Report*, N/T. 41.

The Positive Streamer Mechanism of Spark Breakdown

By W. HOPWOOD*

Department of Electronics, University of Liverpool

ABSTRACT.   The essential conditions for streamer onset in spark breakdown are discussed, with particular reference to the production of photo-ionizing radiation in the initial avalanche.   It is shown that the process of electron–ion recombination can adequately account for the production of the necessary radiation, and is probably the predominant process.   It is also shown that the yield of photons by this mechanism, and consequently the threshold for streamer advance, is critically dependent upon the space-charge field produced by the initial avalanche.   While Meek's criterion in its quantitative form gives the order of this critical space-charge field and enables the breakdown voltage for uniform electric fields to be calculated, the newly proposed conditions for photon production give a new physical significance to the criterion.

§ 1. INTRODUCTION

THE theory of spark breakdown by the positive streamer mechanism is now well known, and the general qualitative picture of streamer propagation by photo-ionization in the gas is well supported by observations.   Measurements of time lags (Loeb 1939) in spark breakdown at atmospheric pressure and the constricted nature of spark channels at high pressures show that the Townsend mechanism involving secondary processes at the cathode is inadequate to explain the observed phenomena.   Observations on positive-point corona (Loeb and Leigh 1937) show the presence of fine streamers originating from the anode, and have led Loeb (1939) to formulate the qualitative theory of streamer propagation by photo-ionization in the gas.

Attempts to determine the quantitative nature of the conditions necessary for streamer propagation were made independently by Meek (Loeb and Meek 1941), and Raether (1939 a, b, 1940).   Both led to the same proposed condition, that the excess field created in the rear of the positive space-charge tip of the avalanche, when approaching the anode, should be of the same order of magnitude as the field applied to the gap.   The calculation of sparking potentials using this criterion, for a number of gap geometries, has been described in detail by Loeb and Meek (1941).

In recent publications, Loeb (1948 a, b), has indicated that the conditions necessary for streamer propagation are : (i) that sufficient high energy photons must be created in the initial electron avalanche to ionize some of the gas atoms or molecules present, (ii) that these photons be absorbed to produce ions in adequate proximity to the streamer tip, and (iii) that the space-charge field at the rear of the avalanche tip shall be great enough to give adequate secondary avalanches in the enhanced field.

* I.C.I. Research Fellow.

In the absence of data on the first two conditions, the Meek criterion was proposed taking account only of the third condition. It was assumed that when condition (iii) is satisfied, (i) and (ii) are automatically satisfied.

On the basis of the above three conditions, Loeb has derived a general equation for the calculation of sparking potentials. The equation is based on the following assumptions:

(*a*) that photo-ionizing radiation is produced in an avalanche by electron collisions with gas molecules, and that the number of photons produced (n_p) is directly proportional to the number of electrons produced (n_e), i.e.

$$n_p = f n_e, \qquad\qquad \dots\dots (1)$$

where *f* is a constant factor, and

(*b*) that a positive streamer will propagate when a single photo-electron, produced in the vicinity of the primary avalanche tip, will create in a secondary avalanche the same total number of electrons, and hence the same number of photons, as were produced in the primary avalanche.

The equation, since it involves the so far uncertain quantities *f* and μ, the absorption coefficient of the gas for the photo-ionizing radiation, cannot be readily applied to the calculation of sparking potentials. A more recent publication (Loeb and Wijsman 1948) describes the derivation of a similar equation assuming that more than one photo-electron will be necessary to advance the streamer tip.

§2. THE MEEK CRITERION

It is the purpose of this work to discuss Loeb's three conditions in the light of the meagre data existing at present, and it is proposed first to consider condition (iii), upon which Meek's criterion is based.

The criterion as first proposed was that the excess field in the rear of the positive space-charge at the tip of the avalanche should be of the same order of magnitude as the applied field. That is

$$X_s = (1 + K) X_g, \qquad\qquad \dots\dots (2)$$

where X_s is the axial field at the rear of the positive space-charge tip, and X_g is the original field applied to the gap. *K* is a constant factor between 0 and 1, initially assumed to be unity.

Further consideration of the criterion, and the sparking potentials derived from Meek's equation, shows that a value for *K* of 0·1 gives a sparking potential of 31·6 kv. for a 1 cm. gap in air at atmospheric pressure; this is in better agreement with the observed value than that of 32·2 kv., calculated assuming $K = 1\cdot 0$.

It has been shown by Zeleny (1942) however, that in order to bring the calculated sparking potentials for long gaps in air into conformity with measured values, one has to assume values of *K* of the order 10^{15}, even though the calculated potentials are only 7% below those observed experimentally.

Since it is obvious from general considerations that *K* is never likely to be greater than 1·0, as Zeleny points out, there appears to be some considerable discrepancy between measured sparking potentials and those calculated from Meek's theory. It can in fact be shown that for avalanches approaching the anode, *K* is never likely to be greater than 0·5, (i.e. $X_s \leqslant 1\cdot 5\, X_g$), since space-charge conditions in the avalanche would then be such as to prohibit any further advance.

The statement of the discrepancy in terms of K, however, gives an exaggerated impression, and as the following consideration of the nature of the calculations shows, one should expect that the calculated sparking potentials will be smaller than the measured values, particularly for longer gap lengths.   In the derivation of Meek's equation, the ion densities at the tip of the avalanche are calculated assuming the Townsend coefficient for ionization by electron collision, corresponding to the applied field X_g, to be operative during the whole time of avalanche transit.   This assumption is clearly not precisely correct, since at potentials near to the sparking potential, when high positive-ion densities are created in the avalanche tip as it approaches the anode, the potential distribution in the gap will be distorted.   The effect of this space-charge distortion will be such that those electrons in the rear of the advancing electron cloud, ahead of the positive space charge, will experience a steadily decreasing field, and consequently the mean effective Townsend coefficient α will be reduced.   Hence ion densities close to the anode will be somewhat smaller than those calculated assuming the Townsend α corresponding to X_g to apply across the whole of the gap, and therefore potentials higher than those calculated from Meek's equation will be necessary to produce sufficient ionization to satisfy Meek's criterion.   Futhermore, the effect should increase at longer gap lengths, since as the gap length is increased X_g/p at sparking decreases, and at the lower X/p where α is small a fairly small change in X reduces α to an almost negligible value, whereas at higher X/p in short gaps, a similar proportional change in X will reduce α by only a few per cent.   It is probable therefore that the discrepancies observed in the application of Meek's equation to large gap lengths can be largely attributed to the use of over simplified assumptions in the derivation of the equation, rather than to a failure of Meek's criterion and the streamer theory of spark breakdown.

Since then there is no outstanding reason for doubting the general validity of the streamer theory, and since the maximum field obtaining in the rear of an avalanche tip is probably $\sim 1 \cdot 5\ X_g$, it is proposed to assume in taking account of condition (iii) that the field contributing to the formation of secondary avalanches will never exceed this value.

§3. ABSORPTION OF PHOTO-IONIZING RADIATION

If one now considers condition (ii), it is clear that a complete assessment of its importance in the streamer mechanism would involve a detailed knowledge of the spectral distribution of radiation emitted from the avalanche, and of the values of absorption coefficients at all wavelengths.   Only in the case of air (Schneider 1940), has a precise experimental determination of absorption coefficients been attempted, and from this it appears that over the large wavelength range observed, the absorption coefficients at atmospheric pressure for photons capable of ionization in the gas are of the order $100\ \mathrm{cm^{-1}}$.   Geballe (1944) has reported a value of $0 \cdot 5\ \mathrm{cm^{-1}}$ in H_2 at a pressure of 1 mm. Hg for radiations produced in a Townsend discharge.   This is also the order of absorption coefficients determined theoretically for a number of atoms (Bates 1939).

These values suggest that any photo-ionizing radiation produced in the avalanche will be effectively absorbed in distances of the order of 1 mm. or less at atmospheric pressure; this is further confirmed by the general filamentary nature of observed streamers.   It is proposed to assume therefore that 1 mm. is the maximum effective absorption distance at atmospheric pressure.

§4. THE CONDITION FOR STREAMER ONSET

One must now consider streamer condition (i) in the light of the approximate magnitudes for the avalanche space-charge field, and the absorption coefficient for the photo-ionizing radiation deduced above.

Loeb (1949 a, b) has proposed the condition that the photons produced in the primary avalanche must themselves produce so many photo-electrons in the vicinity of the avalanche tip that the total number of electrons produced in secondary avalanches will be equal to the total number produced in the primary avalanche. This criterion is based on the assumption that the number of photons produced in the avalanche, of sufficient energy to ionize in the gas, is directly proportional to the number of electrons produced, the photons being produced by electron collisions.

Suppose we now consider the case of a 1 cm. gap in air at atmospheric pressure, where X_g/p is 42·0 volts/cm/mm. Hg. If we assume the field at the rear of the avalanche tip to be $X_s \sim 1·5 X_g$ we have $X_g/p \sim 63$ volts/cm/mm. Hg. Assuming Sanders' (1933) values for the Townsend coefficient α we have $\alpha_g \sim 17·0$ and $\alpha_s \sim 120$. Therefore the number of electrons produced in the primary avalanche is of the order $\exp \alpha_g \sim e^{17}$.

The maximum number of electrons produced in a single secondary avalanche will be of the order $\exp(\alpha_s x)$, where x is the distance from space-charge region at which the photo-electron is absorbed, and taking $x = 0·1$ cm. (§3), $\exp 0·1 \alpha_s = e^{12}$. This is an upper limit, since it is assumed that the field $1·5 X_g$ extends over the whole distance x. Consequently, the number of photons necessary to satisfy the streamer condition will be at least $e^5 \sim 10^2$.

In the case of a 10 cm. gap, $X_g/p \sim 35$ and $X_s/p \sim 53$, so that $\alpha_g \sim 4·7$ and $\alpha_s \sim 57$, again from Sanders' values. The number of electrons produced in the primary avalanche will be of the order of e^{47}, and in a single secondary avalanche will be of the order of e^6. Therefore the number of photons required to replace the primary avalanche electrons, by means of secondary avalanches, will be of the order of 10^{17}.

It is apparent therefore that quite large numbers of high energy photons must be produced in the primary avalanche, in order for the streamer mechanism to proceed, if one assumes that all the primary avalanche electrons must be replaced by secondary processes.

§5. THE PRODUCTION OF PHOTO-IONIZING RADIATION IN THE AVALANCHE

Now the only processes whereby photo-ionizing radiation could be produced by the excitation by electron collision in a pure gas * are (i) ionization from inner shells giving rise to soft x-rays, and (ii) excitation of positive ions to high energy states.

The first of these processes is extremely improbable, and would necessitate very high energies for the colliding electrons. Furthermore, the proportion of

---

* In the case of mixed gases, it is possible that photons produced from excitation by electron collision with one constituent may be of sufficient energy to ionize a constituent of lower ionization potential and thus facilitate streamer propagation. This mechanism is well known and has been discussed by Weissler (1943). The suggestions outlined in the paper are intended to show that streamer propagation, and the success of Meek's criterion in the case of air, can be explained in terms of photon production by the recombination mechanism, and that the mechanism would appear to be the only one applicable to the case of pure gases, where such factors as easily ionizable constituents do not exist.

such photons to ions would be expected to decrease rapidly as X/p decreased, whereas it would appear from the above (§4) that the streamer mechanism requires a higher proportion of photons to electrons at low X/p than at higher field strengths.

The second process of excitation of positive ions by electron impact will now be discussed. Although no precise information on cross sections for this process is available it seems reasonable to assume that the probability of excitation of ions to levels of energy greater than the first ionization potential of the neutral molecule will be of the same order as the probability of ionization of the neutral molecule. It is reasonable to assume therefore that the ratio of ionizing photons to the total number of electrons produced in the avalanche will not be greater than the ratio of positive ion density to neutral molecule density in the dense space-charge region at the tip of the avalanche. This ratio appears to be of the order 10^{-7}, from Meek's calculations.

Therefore in a 1 cm. gap in air at atmospheric pressure where the number of electrons produced in the primary avalanche is of order 10^7, it seems improbable that any photo-ionizing radiation will be produced by this mechanism. In a 10 cm. gap the number of electrons is of order 10^{20}, and therefore the number of photons is of order 10^{13}, which is considerably below the required number.

It seems doubtful, therefore, whether processes of excitation by electron collision are adequate to supply the number of photons necessary for streamer propagation, if it is assumed that the primary avalanche electrons are lost to the anode, and must be completely replaced in secondary avalanches.

§6. THE RECOMBINATION PROCESS

A third mechanism by means of which high energy photons could be produced, but which has not so far been seriously considered, is that of electron–positive-ion recombination.

If one considers again the case of the 1 cm. gap in air at atmospheric pressure, the maximum rate of production of photons by recombination in the whole avalanche tip will be given by

$$\frac{dn_p}{dt} = \alpha N_+ n_-, \qquad \qquad \ldots\ldots(3)$$

where α is the coefficient of recombination to the ground-state, which will supply most of the photons capable of ionization, N_+ is the density of positive ions at the tip of the avalanche, and n_- is the total number of electrons produced in the avalanche.

Now if we assume $\alpha = 10^{-10}\,\text{cm}^3/\text{sec}$. which is a value determined for the F_2 region of the upper atmosphere (Sayers 1942–43), and is probably a reasonable upper limit for radiative capture (Massey and Bates 1942–43), we have

$$dn_p/dt = 10^{-10} \times 3 \times 10^{12} \times 2 \times 10^7/\text{sec}. \sim 10^{10}/\text{sec}.$$

since $N_+ \simeq 3 \times 10^{12}$ and $n_- \simeq 2 \times 10^7$ for a 1 cm. gap, according to the calculations of Meek (Loeb and Meek 1941).

If we assume that conditions in the gap are not influenced by space charge, the drift velocity of the electrons is $2 \times 10^7\,\text{cm/sec}$., and since the majority of the positive ions are formed in the final 1 mm. of the gap, the time of transit of the electrons through this dense space-charge region will be $\sim 0.5 \times 10^{-8}\,\text{sec}$.

Therefore the total number of photons produced by recombination before the electrons are swept into the anode is probable not greater than 50, and since the photons are emitted in all directions, it is probable that no more than a quarter of them (say 10 photons), will be absorbed in the region behind the avalanche tip, where they could effectively contribute to the streamer mechanism. These 10 operative photons are smaller in number than the 100 required (see §4 above), but the deficit may easily be made up by a reduction in electron velocity, as is explained below.

In the case of the 10 cm. gap the number of photons emitted is of the order 10^{13}, which is again considerably smaller than the 10^{17} required by the Loeb streamer condition. It would appear, therefore, that under the assumed conditions, the mechanism of electron–positive-ion recombination is also inadequate to supply the photons necessary for streamer formation.

The assumed conditions are, however, in error, since the space charges produced will influence the progress of the avalanche. In fact as the avalanche is approaching the anode it is leaving behind a considerable positive space charge which will, so far as the advancing electrons are concerned, tend to neutralize the applied gap field. Consequently, the velocity of those electrons in the rear of the advancing electron cloud will be reduced, and the effect will be a decrease in the effective Townsend coefficient, as has been pointed out previously (§2) with regard to the calculated sparking potentials for long gaps. Furthermore, the transit time of the electrons through the dense space-charge region will be increased, and this will tend to increase the yield of photons produced by the process of electron–ion recombination. Also the recombination process will be further enhanced in preference to processes of excitation, as the reduced electron energy will result in an increase in recombination coefficient, whereas processes of excitation will be increasingly improbable.

It is probable, therefore, that the condition necessary for a sufficient yield of photo-ionizing radiations to be produced in the avalanche is that the space-charge fields produced shall be of such magnitude as to affect the progress of the avalanche. It is possible, particularly in the case of longer gaps, where a higher rate of photon production is required, that this condition will have to be taken to the limit, and sufficiently intense fields will be required to hold back a considerable proportion of the electrons in the positive space-charge region, when something approaching a plasma will be produced.

The magnitude of the necessary fields will obviously be of the same order as those demanded by Meek's criterion in its quantitative form. The physical significance to be attached to the criterion, however, is different from that previously assumed, and more emphasis should be attached to the retarding effects of the positive space charge on the advancing electrons than to its effect on the production of secondary avalanches. It would appear to follow as a natural consequence of the conditions necessary for adequate photon production, that sufficiently intense fields will be present in the rear of the primary avalanche for adequate electron multiplication in secondary avalanches.

When the applied gap field is of sufficient magnitude to bring about the required space-charge condition, and a number of electrons are trapped in the dense positive space-charge region at the anode, the streamer mechanism will proceed, if the photon production is sufficient to replace the electrons in the space-charge

region, by secondary avalanches, at a rate equal to or greater than that at which they are being lost to the anode. This necessary rate of photon production, however, will not be so great in the case where electrons are to some degree bound to the positive space charge, as that required by Loeb's criterion, where it is assumed that all the primary avalanche electrons are lost to the anode, and must be completely replaced by secondary avalanches.

As the streamer advances into the gap, it will probably extend itself as a low field space-charge column, with a region of high charge density at the tip. In order for the streamer to propagate it will be only necessary for electrons to feed into the tip at the same rate as they are being drawn from the tip, up the column to the anode. Because of the low fields probably existing in the streamer channel, the rate of electron loss will not be very high.

When the desired space-charge condition is attained at the anode in longer gaps, of 10 cm. or more, it may happen that the rate of photon production will not be sufficient to compensate for the loss of electrons to the anode, since the ion densities attained are not so high as in shorter gaps, and the Townsend α effective in secondary avalanche production has only a low value. Under these conditions, it will be necessary to apply higher potentials to produce breakdown, in which case the desired space-charge condition will be achieved somewhere in mid-gap, with an increase of ion density and field strength responsible for secondary avalanche production. With increasing applied gap field the space-charge effect will choke off the avalanche at increasing distances from the anode, until the rate of photon production becomes sufficiently high to give a return positive streamer to the cathode, when breakdown may take place by the avalanche retrograde streamer mechanism proposed by Loeb (Loeb and Meek 1941).

§ 7. CONCLUSION

It is probable that the process of electron–positive-ion recombination is the predominant process in the production of photo-ionizing radiation in the initial stages of spark breakdown. Meek's criterion gives the order of space-charge field necessary for sufficient photon production by this mechanism, and adequate multiplication of electrons in secondary avalanches follows as a natural consequence of this space-charge condition. Deviations from measured values in sparking potentials for longer gaps calculated from Meek's equation can be attributed to over-simplified assumptions used in the derivation of the equation.

Any accurate calculations of sparking potentials, using the Meek criterion, or the modified form of it suggested above, must take into account space-charge effects occurring during the electron avalanche transit, and the corresponding change in Townsend's first coefficient, caused by these effects. It would be necessary in attempting such calculations to have a knowledge of the degree of lateral spreading of the electron avalanche during its passage between the electrodes. Data on this spreading are very limited, and since it probably arises from a complex diffusion process involving clouds of electrons and positive ions in close proximity, an experimental approach is likely to yield more useful information than a theoretical treatment. It is proposed to attempt experiments in this laboratory on the determination of electron avalanche growth. Investigations on the absorption of photo-ionizing radiation in a number of gases by spectroscopic methods are now in progress.

ACKNOWLEDGMENTS

The author is grateful to Professor J. M. Meek, in whose Department this work was carried out, and to Dr. J. D. Craggs for helpful criticism and discussions on some aspects of the work. Thanks are also due to Professor J. Sayers of Birmingham University, for helpful advice, and for pointing out a number of errors in the text, and to Professor L. B. Loeb for comments.

REFERENCES

BATES, D. R., 1939, *Mon. Not. R. Astr. Soc.*, **100**, 25.
GEBALLE, R., 1944, *Phys. Rev.*, **66**, 316.
LOEB, L. B., 1939, *Fundamental Processes of the Electrical Discharge in Gases* (New York and London : J. Wiley and Sons) ; 1948 a, *Rev. Mod. Phys.*, **20**, 151 ; 1948 b, *Phys. Rev.*, **74**, 210.
LOEB, L. B., and KIP, A. F., 1939, *J. Appl. Phys.*, **10**, 142.
LOEB, L. B., and LEIGH, W., 1937, *Phys. Rev.*, **51**, 149 (A).
LOEB, L. B., and MEEK, J. M., 1941, *Mechanism of the Electric Spark* (Stanford : University Press).
LOEB, L. B., and WIJSMAN, R. A., 1948, *J. Appl. Phys.*, **19**, 797 (A).
MASSEY, H. S. W., and BATES, D. R., 1942–3, *Rep. Prog. Phys.*, **9**, 62 (London: Physical Society).
RAETHER, H., 1939 a, *Z. Phys.*, **112**, 464 ; 1939 b, *Ibid.*, **117**, 375 and 524 ; 1940, *Arch. Elektrotech.*, **34**, 49.
SANDERS, F. H., 1933, *Phys. Rev.*, **44**, 1020.
SAYERS, J., 1942–3, *Rep. Prog. Phys.*, **9**, 52 (London : Physical Society).
SCHNEIDER, E. A., 1940, *J. Opt. Soc. Amer.*, **30**, 128.
WEISSLER, G. L., 1943, *Phys. Rev.*, **63**, 96.
ZELENY, J., 1942, *J. Appl. Phys.*, **13**, 444.

REVIEWS OF BOOKS

Physical Aspects of Colour, by P. J. BOUMA. Pp. 312. Translation from the Dutch original. (Eindhoven : Philips Industries, 1947 ; British Agents: Cleaver-Hume Press.) 30*s*.

This textbook is indeed welcome coming from a continental author. We have become so accustomed to the Anglo-American (or should it be American-Anglo) point of view in colour matters, that it is quite a refreshing change to have the spot-light tilted more towards the work of European investigators. Although the title is *Physical Aspects of Colour* and the preface asserts that the book is for " all persons interested in the origin and the measurement of colours", yet little attention is given to the physical analysis of radiation itself. For example, the concepts of radiation energy and light absorption are used freely, yet the term " thermopile " is never mentioned and the whole of practical spectrophotometry is dismissed in three pages. It is to be assumed therefore that the author's real intention is to take the physical side with its technique of measurement more or less for granted and to go on from there. In compensation for this we have a wide-ranging exposition of the psycho side of psycho-physical colour. The author is widely read in this field and the appended bibliography is full and well laid out. The historical aspect of the subject, generally neglected or barely mentioned in the familiar textbooks, is treated here fully and interestingly. It is certainly convenient for example to have the gist of Goethe's teaching presented in English and appraised by a practical physicist.

The author begins, rightly, with a discussion on brightness, in the new terminology, luminance. The ideas connected with equality of luminance are in some ways more difficult to handle than those of simple colour matching. The luminosity function of the eye, however, was standardized first and it has so influenced the procedure in setting up

the colour functions that a natural presentation of the C.I.E. system is impossible without a preliminary discussion of the luminosity curve. In this instance, the presentation is not made easier by introducing measurement in the region of varying luminosity curve under the heading " subjective brightness " even before photopic brightness has been defined.

The general theory of additive colour mixture based on Grassmann's laws is set forth in a very austere fashion. The characteristics of colour space follow with a discussion of the projective transformation and the C.I.E. system. Dr. Bouma professes to be an adherent of colour-space representation but it is perhaps significant that subsequent to the particular chapter on colour space some forty chromaticity diagrams are used to illustrate the text but only four diagrams of colour space.

The discussion of brightness and colour matching occupies about half of the book. The remainder is devoted to a wide variety of topics ranging from Temperature Radiation to, say, " Characteristic Colours " (characteristic colours are defined as the most colourful examples among the ideal colours).

Taking the book as a whole, a lack of adequate discussion is often felt concerning the experimental foundations underneath the profusion of ideas and results. To the enquiring reader the unanswered question " What was really measured and how was it done ? " frequently arises. This book would be rather difficult as a first textbook for a beginner : its virtues can be better appreciated by the more experienced.   R. D.

Atmospheric Electricity, by J. ALAN CHALMERS. Pp. 175. First Edition. (Oxford : University Press (Geoffrey Cumberlege), 1949). 15s.

The contributions of British workers to the subject of atmospheric electricity have been considerable, with C. T. R. Wilson, G. C. Simpson, F. J. Whipple, and their collaborators coming to mind as fairly recent leaders in the field. Yet the only texts to have been published in this country up to 1948 were Schonland's small though very attractive monograph of 1932, now long out of print, and a translation from the German of Hess on atmospheric conductivity. That position is to be contrasted with a series of texts which appeared in Germany and France during the inter-war years, and a fairly full but by no means complete treatment in *Terrestrial Magnetism and Electricity*, published in 1939 in the United States. How disappointing then to find that when, at long last, a treatment, comprehensive in intent, appears in this country it should be so far from satisfying.

Dr. Chalmers intends to provide " an introduction for the uninitiated . . . . and a work of reference for workers on the subject ". " Written primarily for physicists, it should yet prove useful ", so the dust cover informs us, " to meteorologists and others interested in the subject. " (How are the meteorologists flattered !) This multiplicity of aim is probably the author's undoing. He will, quite certainly, fail to satisfy the worker in the subject and, in the reviewer's opinion, the uninitiated would do better to go back to Schonland or to C. T. R. Wilson's article in the *Dictionary of Applied Physics*, Vol. 3.

The main criticism is of jerkiness and lack of form in a treatment which might have been expected to achieve continuity and structure. For atmospheric electricity, in contrast with some branches of geophysics, knows well what it is about with aims already partly realized or, as it appears, eminently approachable. These aims are to provide an explanation of the sign, general level and variation of the earth's fine-weather field and of the fields of disturbed weather, to relate the former to the latter where possible, and to determine the physics of the individual phenomena in relation to weather processes. The choice of the field as the characteristic of the electrical condition of the atmosphere is arbitrary and others might therefore choose to state the aims in somewhat different terms, without however showing any real conflict of aim.

The author appears to have been nervous of his material. Thus, though there are few aspects of the subject which do not receive notice at the head of a paragraph, observations are in general scantily reproduced and ineffectively presented, methods of investigation often receive only slight mention, and a discussion of results as often as not leaves off almost as soon as the reader's attention is engaged. In a designedly comprehensive text (which incidentally would occupy more than 170 of these pages for its realization) one

might have expected to find some mention of the more recent theories which have been advanced to account for the charge separation processes in clouds, particularly in thunderclouds, for example those of Frenkel and Findeisen. Few will agree with the author's statement that " the general processes at work in the thundercloud appear to be fairly well undertsood and the theories are in a reasonably satisfactory state." No theory is satisfactory which fails to give even an order of magnitude for the quantity whose existence it seeks to explain, and, Frenkel apart, that is the present state of theory of the charge separation in clouds. P.A.S.

Outlines of Physical Chemistry, by FARRINGTON DANIELS. Pp. viii + 713. First Edition. (J. Wiley and Sons, Inc., New York, 1948; British Agents: Chapman and Hall, Ltd., London.) $5.

Both teachers and students of Physical Chemistry over a long period have regarded highly the successive editions of Getman's *Outlines of Theoretical Chemistry*, first published in 1913, and carried on by Professor Daniels from 1931. The present volume is an entirely new book of more advanced standard, but it may be regarded as the logical successor to the earlier series, and it inherits many of the admirable qualities of the previous *Outlines*. In particular, the presentation is very readable and lucid without in any way losing accuracy and precision. Following a general introduction, the physical chemistry of the gaseous state is first discussed, and this is followed by a section on interatomic and intermolecular forces, leading to the consideration of the crystalline state and the correlation of physical properties and molecular structure. Thermodynamics and thermochemistry are then introduced, preceding the consideration of the liquid state and the properties of solutions. Then follows the discussion of equilibria in homogeneous and heterogeneous systems, and a very full consideration of chemical kinetics and mechanism of reaction from the modern viewpoint. The electrical properties of solutions and ionic equilibria are then treated, followed by sections on the colloidal state, photochemistry and quantum theory, and in the final chapter modern work on atomic and nuclear structure is described. A very useful appendix contains detailed mathematical deductions which have been omitted from the text, and helpful collections of formulae and constants. To each chapter there is appended a collection of numerical problems, and a list of selected references for further reading. The work can be unreservedly recommended as a stimulating and modern textbook which largely meets the needs of physical chemistry students up to their final honours year, when more specialized works must be consulted. Format and printing are excellent and the text is remarkably free from errors. A. H.

RECENT PUBLICATIONS RECEIVED

How to Study Physics, by SEVILLE CHAPMAN. Pp. iv + 28. Revised Edition. (Cambridge, Mass.: Addison-Wesley Press Inc., 1949.) 25 cents.

Die Brechzahlen einiger Halogenidkristalle, by H. HARTING. Pp. 25. Sitzungsberichte der Deutschen Akademie der Wissenschaften zu Berlin, Mathematische-Naturwissenschaftliche Klasse, Jahrgang 1948, Nr. IV. (Berlin: Akademie-Verlag, 1948,) 2.50 D.M.

Max Planck: In Seinen Akademie-Ansprachen (Erinnerungs-schrift der Deutschen Akademie der Wissenschaften zu Berlin). (Berlin: Akademie-Verlag, 1948.)

CONTENTS FOR SECTION A

ABSTRACTS FOR SECTION A

On the Ratio of Positive to Negative Particles in the Vertical Cosmic-Ray Beam at Sea Level, by B. G. Owen and J. G. Wilson.

ABSTRACT. A method of direct momentum measurement has been used to determine the ratio of positive to negative particles in the cosmic-ray beam at sea level. Measurements on about 12,000 particles give the ratio $+/-=1 \cdot 268 \pm 0 \cdot 023$, and it is further shown that to the statistical accuracy of the experiment there is no variation of this ratio with momentum over the momentum range 10^9–10^{10} ev/c.

The Effect of Holes in a Reacting Material on the Passage of Neutrons, by D. J. Behrens.

ABSTRACT. In this paper we discuss the increase in the migration length of neutrons in a reactor caused by the presence of holes in the reactor. It is shown how this effect can be evaluated in terms of a simple geometric function of the shape of the holes, and it is found that, in addition to a variation as the inverse overall density of the material in the reactor, the migration length also contains a term which depends upon the size and shape of each hole. When the hydraulic radius of the hole becomes comparable to the mean free path of neutrons in the reactor material, this last term becomes dominant, and it is therefore desirable that all holes should be kept as small as possible. Holes which are totally enclosed, or which contain a uni-directional infinity (cylinders) all give results of the same order of magnitude, but anything in the nature of a two-dimensional plane gash leads to a theoretically infinite increase in the migration length, and hence to a serious escape of spare neutrons.

We have considered the cases both of holes of regular shape and of the interstices between a random arrangement of regular-shaped solid bodies.

Betatron Injection into Synchrotrons, by F. K. GOWARD.

ABSTRACT. The paper states simply the factors influencing the proportion of electrons which may be trapped and accelerated in a transition from betatron to synchrotron operation; this transition is a vital factor influencing the design of the normal electron-synchrotron. Machines of various energies are studied, and it is shown that the important considerations in the design are quite different in different energy ranges. The theoretical predictions are compared with experimental evidence obtained on 14 and 70 mev. synchrotrons, and the agreement is sufficient to give confidence in the simplified theory.

A Geiger Counter Spectrometer for the Measurement of Debye–Scherrer Line Shapes, by W. H. HALL, U. W. ARNDT and R. A. SMITH.

ABSTRACT. A spectrometer for the accurate measurement of x-ray powder diffraction line intensities and shapes is described. Monochromatic radiation is used, and the diffracted beam intensity is recorded by a Geiger counter. The intensity of the x-ray source is monitored continuously, and errors due to variations in its output are automatically compensated. The accuracy and resolution obtainable are demonstrated by some typical experimental results.

Surface Forces in Liquids and Solids, by C. GURNEY.

ABSTRACT. The author's recent treatment of the forces in the surfaces of liquids is extended to the surfaces of solids and to interfaces between solids and liquids. Solid surfaces formed by cleavage at temperatures such that no appreciable migration of molecules takes place, will usually be under a state of uniform stress, not necessarily tensile and not numerically equal to the free energy of the surface. If appreciable atomic migration takes place, the surface of a crystalline solid melts, and the solid is covered with a thin liquid film under a tension force greater than that of the corresponding supercooled liquid, and such that the chemical potential of the molecules in the liquid film is equal to that of molecules in the crystal. This tension force is numerically equal to the free energy of the surface. If such a solid is subsequently cooled to a temperature at which atomic migration effectively ceases, it will have frozen in its surface a tension force corresponding to thermal equilibrium at some higher temperature. The Laplace Q force is discussed in terms of modern theories of atomic bonding and applied to the case of equilibrium at angle of contact between solid and liquid.

The Hydrodynamics of Helium II, by R. B. DINGLE.

ABSTRACT. In this paper expressions are derived for the forces acting on the normal fluid and on the superfluid of helium II, and it is shown that the apparently differing derivations of the velocity of second sound given by Landau and Tisza are, in fact, equivalent, even if thermal expansion is not neglected.

The Dependence of the Thermal Diffusion Factor on Temperature, by K. E. GREW.

ABSTRACT. The thermal diffusion factor for isotopic mixtures has been calculated from values of the collision integrals given by Hirschfelder, Bird and Spotz which refer to molecules exerting an inverse 13th power repulsion with an inverse 7th power attraction—the (13, 7) model. The theoretical results are compared with experimental ones for the inert gas mixtures previously considered by Grew in relation to the (9, 5) model. The agreement is, as expected, better than for the (9, 5) model. Experimental results for hydrogen–helium mixtures are also given; the behaviour of these mixtures is exceptional in that the thermal diffusion factor appears to be independent of temperature over the range 12 to 600° K.

PHYSICAL SOCIETY PUBLICATIONS

Fellows and Student Members of the Society may obtain ONE copy of each publication at the price shown in brackets. In most cases the cost of postage and packing is extra.

Resonant Absorbers and Reverberation. Report of the 1947 Summer Symposium of the Acoustics Group of the Physical Society. Pp. 57. In paper covers. 7s. 6d. (5s.) Postage 6d.

The Emission Spectra of the Night Sky and Aurorae, 1948. Papers read at an International Conference held under the auspices of the Gassiot Committee in London in July 1947. Pp. 140. In paper covers. 20s. (12s. 6d.) Postage 6d.

The Strength of Solids, 1948. Report of Conference held at Bristol in July 1947. Pp. 162. In paper covers. 25s. (15s. 6d.) Postage 8d.

Report of International Conference on Fundamental Particles (Vol. I) *and Low Temperatures* (Vol. II), 1947. Conference held at Cambridge in July 1946. Pp. 200 (Vol. I), pp. 184 (Vol. II). In paper covers. 15s. each vol. (7s. 6d.) Postage 8d.

Meteorological Factors in Radio-Wave Propagation, 1947. Report of Conference held jointly with the Royal Meteorological Society in April 1946. Pp. 325. In paper covers. 24s. (12s. + postage 1s.)

Handbook of the 33rd Exhibition of Scientific Instruments and Apparatus, 1949. Pp. 272. In paper covers. 5s. (2s. 6d.) Postage 1s.

Catalogue of the 32nd Exhibition of Scientific Instruments and Apparatus, 1948. Pp. 288. In paper covers. 5s. (2s. 6d.) Postage 1s. (Half price from 5th April 1949.)

Catalogue of the 31st Exhibition of Scientific Instruments and Apparatus, 1947. Pp. 298. In paper covers. 2s. 6d. (1s. 6d.) Postage 1s.

Report on Colour Terminology, by a Committee of the Colour Group. Pp. 56. In paper covers. 7s. (3s. 6d.)

Report on Defective Colour Vision in Industry, by a Committee of the Colour Group. 1946. Pp. 52. In paper covers. 3s. 6d. (1s. 9d. + postage 4d.)

Science and Human Welfare. Conference held by the Association of Scientific Workers, Physical Society and other bodies. 1946. Pp. 71. In paper covers. 1s. 6d. (9d.) Postage 4d.

Report on the Teaching of Geometrical Optics, 1934. Pp. 86. In paper covers. 6s. 3d. Postage 6d.

Report on Band Spectra of Diatomic Molecules, 1932. By W. JEVONS, D.Sc., Ph.D. Pp. 308. In paper covers, 25s.; bound in cloth, 30s. (15s.) Postage 1s.

Discussion on Vision, 1932. Pp. 327. In paper covers. 6s. 6d. (3s. 3d.) Postage 1s.

Discussion on Audition, 1931. Pp. 151. In paper covers. 4s. (2s.) Postage 1s.

Discussion on Photo-electric Cells and their Application, 1930. Pp. 236. In paper covers. 6s. 6d. (3s. 3d.) Postage 8d.

The Decimal Bibliographic Classification (Optics, Light and Cognate Subjects), 1926. By A. F. C. POLLARD, D.Sc. Pp. 109. Bound in cloth. 4s. (2s.) Postage 8d.

Motor Headlights, 1922. Pp. 39. In paper covers. 1s. 6d. (9d.) Postage 4d.

Report on Series in Line Spectra, 1922. By A. FOWLER, C.B.E., Sc.D., F.R.S. Pp. 182. In paper covers. 30s. (15s.) Postage 8d.

A Discussion on the Making of Reflecting Surfaces, 1920. Pp. 44. In paper covers. 2s. 6d. (1s. 3d.) Postage 4d.

Reports on Progress in Physics. Vol. XI (1946–48). Pp. 461. Bound in cloth. 42s. (25s.) Postage 1s.

Reports on Progress in Physics. Vols. IV (1937, reprinted 1946) and X (1944–45). Bound in cloth. 30s. each. (15s.) Postage 1s.

The Proceedings of the Physical Society. From Vol. I (1874–75), excepting a few parts which are out of print. Prices on application.

The Transactions of the Optical Society. Vols. 1 (1899–1900) –33 (1931–32), excepting a few parts which are out of print. Prices on application.

Orders, accompanied by remittances, should be sent to

THE PHYSICAL SOCIETY

1 Lowther Gardens, Prince Consort Road, London S.W. 7